University of Alberta

Modeling and Monitoring of a High Pressure Polymerization Process using Multivariate Statistical Techniques

by

Rumana Sharmin

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Process Control

Department of Chemical and Materials Engineering

Edmonton, Alberta

Fall 2007

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-33063-0
Our file Notre référence
ISBN: 978-0-494-33063-0

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Read! In the name of thy Lord Who created- Created man, out of a (mere) clot of congealed blood; Read: And thy Lord is the Most Bountiful, He Who taught (the use of) the pen,- Taught man that which he knew not.

The Holy Quran (96:1-5)

Dedicated to:

My parents

Dr. Hosne Ara Begum & Dr. Mohammad Habibur Rahman

Abstract

This thesis explores the use of multivariate statistical techniques in developing tools for property modeling and monitoring of a high pressure ethylene polymerization process. In polymer industry, many researchers have shown, mainly in simulation studies, the potential of multivariate statistical methods in identification and control of polymerization process. However, very few, if any, of these strategies have been implemented. This work was done using data collected from a commercial high pressure LDPE/EVA reactor located at AT Plastics, Edmonton. The models or methods developed in the course of this research have been validated with real data and in most cases, implemented in real time.

One main objective of this PhD project was to develop and implement a data based inferential sensor to estimate the melt flow index of LDPE and EVA resins using regularly measured process variables. Steady state PLS method was used to develop the soft sensor model. A detailed description of the data preprocessing steps are given that should be followed in the analysis of industrial data. Models developed for two of the most frequently produced polymer grades at AT Plastics have been implemented. The models were tested for many sets of data and showed acceptable performance when applied with an online bias updating scheme.

One observation from many validation exercises was that the model prediction becomes poorer with time as operators use new process conditions in the plant to produce the same resin with the same specification. During the implementation of the soft sensors, we suggested a simple bias update scheme as a remedy to this problem. An alternative and more rigorous approach is to recursively update the model with new data, which is also more suitable to handle grade transition. Two existing recursive PLS methods, one based on NIPALS algorithm and the other based on kernel algorithm were reviewed. In addition, we proposed a novel RPLS algorithm which is based on the Krylov Controllability based PLS theory. It was found that the new method is much faster, and hence is suitable for time varying system containing many variables.

Finally, we present a data based monitoring scheme to detect the onset of decomposition in a LDPE reactor. This novel method combines PCA and an energy balance around the reactor. This relatively simple method was able to detect the onset of decomposition with reasonable lead time. The method has been implemented at the plant where it is being used as an additional monitoring tool to ensure safe reactor operation.

Acknowledgements

I express my deepest gratitude to my supervisors Dr. Sirish L. Shah and Dr. Uttandaraman Sundararaj for their excellent guidance and careful supervision throughout the course of my PhD research work. I most respectfully acknowledge their timely advice and encouragement that enabled me to complete successfully in time. My words are insufficient to express my gratefulness towards both of them for their endless support.

I am thankful to Dr. Larry Vande Griend and Dr. Yi Jun Sun for providing me with the technical support at AT Plastics during last five years. They were very kind to explain and answer my many naïve questions about the industrial process. I am extremely grateful to my friend Ian Allyene for his help in data collection and model implementation at the plant. From the beginning till the completion of my work at AT Plastics, Ian was one of my most valuable sources of support for his excellent skills in computers and his unlimited patience. I am also indebted to Dirk Hair, Colin Sikora, Dave Rozak, Jerald, Emil, Brent, Bavin, and all the plant operators at the AT Plastics control room for their help at numerous occasions.

I would like to thank Alberta Ingenuity Fund for providing me with the financial support through the Alberta Ingenuity PhD scholarship. I also thank the NSERC-Matrikon-Suncor-iCORE Industrial Research Chair program and AT Plastics for giving me the additional financial support.

It was a very memorable experience for me to be a member of the process control group at the University of Alberta. I thank all my friends and colleagues for creating a friendly and cooperative workplace. I express my sincere gratitude to Dr. Arun K Tangirala, and Dr. Weihua Li for their valuable comments and suggestions throughout my project. My sincere gratitude goes to Dr. Biao Huang for his excellent teaching in a number of advanced process control courses. I had many useful discussions with my colleagues

Zhengang, Hari, Vikas, Fangwei, Syed, Shoukat, Salim, Manjur, Ian, Liqian, Adrian, Jana, and David. A special thank is reserved for Vikas for introducing me to the people at AT Plastics. I also thank Dr. Shankar Narasimhan and Dr. Kannan Moudgalya for their valuable suggestions.

I am grateful to the technical stuffs at the department, particularly Bob Barton, Jack Gibeau and Greg Gibeau for their support with computers and networking. I thank AnnMarie, Theressa, Marion, and Leanne for helping me with numerous official issues.

I express my sincerest gratitude to my parents for their love and inspiration all through my life. My deepest thank to my husband Dr. Tariq M. Mannan for always being by my side, giving me inspiration and mental support, for his patience and cooperation. This work would not have been possible without his help. Very special thanks go to my two loving sons Tawfeeq and Tauqir who were born during the course of this PhD project. I am extremely grateful for all the sacrifice they had to make to make my work successful.

And finally, I thank Almighty Allah for all His blessings.

Contents

1 Introduction 1

1.1 Introduction (Overview of Global Polymer Market) 1

1.2 Polyethylene: Manufacturing Processes and the Product 2

1.3 High Pressure LDPE/EVA Process Technology 6

1.4 Modeling and Control of Polymerization Processes 7

1.5 Introduction to Multivariate Statistical Methods 10

1.6 Organization of the Thesis 14

2 Review: Polymer Property Modeling 16

2.1 Introduction 16

2.2 Polymer and Polymerization Processes 17

2.3 Quality Control in Polymer Industry 22

2.3.1 First principles models 27

2.3.2 Black Box Models: Use of Neural Networks 32

2.3.3 Data based Statistical Models 35

2.4 Summary: Challenges & Motivation 38

3 Data Based Inferential Modeling 41

3.1 Introduction 41

3.2 Partial Least Squares: Theory 42

3.3 Process Description 48

3.4 Grade Transition and Modeling strategy 51

3.5 Data Acquisition and Preprocessing 53

3.5.1 Data Collection 53

3.5.2 Assessing Data Quality 55

3.5.3 Sorting for In-line Pumps 61

3.5.4 Calculation of mass flow rate 62

3.5.5 Time delay estimation 62
3.6 Results from PLS Models 63
3.7 Model implementation: Need for updating 68
3.8 Summary 79

4 Recursive PLS for Time Varying Process **80**
4.1 Introduction 80
4.2 Recursive PLS Theories: an Overview 81
4.3 RPLS based on NIPALS algorithm 85
4.4 RPLS based on Kernel Algorithm 90
4.5 KCM based PLS theory and RPLS Extensions 97
4.5.1 Eigenvalue, Eigenvector, Caley Hamilton theorem 98
4.5.2 Krylov/Controllability Matrix 99
4.5.3 Krylov controllability matrix based PLS algorithm 99
4.5.4 Optimal Estimation of α^*_{lo} 102
4.5.5 Modified KCM based PLS theory and RPLS 104
4.6 Grade transition, Data Quality and artifacts of MI Reading 109
4.7 Modeling grade transition 112
4.8 Implementation of the RPLS algorithms 117
4.8.1 NIPALS based RPLS method: results 117
4.8.2 Kernel based RPLS method: Results 124
4.8.3 KCM based RPLS method: Results 127
4.9 Summary 130

5 PCA Based Reactor Monitoring **131**
5.1 Introduction 131
5.2 Decomposition in High Pressure LDPE Rectors 133
5.3 Principal Component Analysis: Theory 135
5.3.1 Fault detection using PCA 138
5.3.2 PCA based Model Identification 142
5.4 Process description 146

5.5 Formulation of Overall Heat Balance around the reactor 151
5.6 Order of Magnitude Analysis 153
5.7 PCA Model Development 157
5.8 Choice of Appropriate PC 159
5.9 Results and Discussion 175
5.10 Summary 191

6 Concluding Remarks **192**
6.1 Summary 192
6.2 Contributions of this thesis 194
6.3 Recommendations for future work 195

Bibliography **198**
Appendix **209**

List of Tables

Table 1-1	High pressure process versus low pressure process	4
Table 1-2	Comparative summary of different Polyethylene resins	5
Table 2-1	Polymerization mechanism in step and chain polymerizations	19
Table 2-2	Relationship between molecular and physical properties of polymer (adapted from Ohshima and Tanigaki, 2000)	25
Table 2-3	Summary of literature on high pressure LDPE reactor modeling	29
Table 2-4	Summary of literature on Neural Network based modeling of polymerization	33
Table 3-1	Sample of the data as collected from the plant site	54
Table 3-2	Estimated Transport Delay	62
Table 4-1	Batch wise NIPALS algorithm for PLS	86
Table 4-2	NIPALS based block wise RPLS algorithm	89
Table 4-3	Kernel Algorithm for PLS	93
Table 4-4	Modified Kernel algorithm for PLS	96
Table 4-5	Sample wise RPLS based on modified kernel algorithm	97
Table 4-6	Summary of Krylov controllability matrix (KCM) based PLS algorithm	104
Table 4-7	Modified Krylov controllability matrix (KCM) based PLS algorithm.	107
Table 4-8	Summary of NIPALS based RPLS Model Performance	123
Table 4-9	Summary of Kernel based RPLS Model Performance	126
Table 4-10	Summary of Controllability Matrix based RPLS Model Performance.	129
Table 5-1	PCA based model identification using noisy data	146
Table 5-2	Calculation of various terms in X matrix	160
Table 5-3	Results from the normal PCA Model	164
Table 5-4	Loading vectors from the normal PCA Model	164

List of Figures

Figure 2-1	(a) Polymerization of ethylene into Polyethylene; (b) Structure of Ethylene Vinyl Acetate copolymer obtained from polymerization of ethylene and vinyl acetate	17
Figure 2-2	Linear and Branched Polymer	22
Figure 2-3	Relationships among polymer properties and processing condition	24
Figure 3-1	PLS Block Diagram	44
Figure 3-2	Polymerization process flow diagram	51
Figure 3-3	Illustration of swinging door compression. Black circles represent archived spot values. Values with open circles are not archived. At time step 5 (point y_e) the lower door (dotted line) opens up wider than parallel shows that a new trend started at y_d	56
Figure 3-4	Plot of fresh feed flow rate shows how the data is stored in the historian	58
Figure 3-5	Example of data quantization	59
Figure 3-6	Prediction from high MI PLS model	64
Figure 3-7	Prediction from low MI PLS model	65
Figure 3-8	PLS model prediction for grade-D	66
Figure 3-9	PLS model prediction for grade-C	67
Figure 3-10	PLS model prediction for grade-E	68
Figure 3-11(a)	Model validation with three new data sets for Grade-D (after 3 months)	69
Figure 3-11(b)	Model validation with three new data sets for Grade-D (after 4 months)	70
Figure 3-11(c)	Model validation with three new data sets for Grade-D (after 15 months)	71
Figure 3-12(a)	Variation in reactor pressure	71

Figure 3-12(b)	Variation in reactor feed gas temperature	73
Figure 3-13	Model validation new data set for Grade-D: a) scaling factors taken from training data, b) scaling factor taken from validation data	74
Figure 3-14	Model performance for Grade D after 4, 5, 13, 14, and 15 months (from top to bottom) without update (Left); with bias update at every hour (Right); (solid line: true MI, dashed line: model prediction)	76
Figure 3-15	Prediction versus Lab MI samples: a) without bias update and b) with bias update	77
Figure 3-16(a)	Online implementation result for softsensor developed for grade D (July 2006)	78
Figure 3-16(b)	Online implementation result for softsensor developed for grade D (Nov 2006)	79
Figure 4-1	Melt Index Change during a Grade Transition from 6 to 4 MI.....	110
Figure 4-2	Online Melt Index during a grade change from 6 to 445 MI	111
Figure 4-3	Change in product properties during grade transition; Top: change in melt index; Bottom: Change in Vinyl Acetate content..	113
Figure 4-4	Changes in reactor pressure (Top) and reactor temperature (Bottom) during grade transition	114
Figure 4-5	Changes in propylene (Top) and VA mass flow rate (Bottom) during grade transition	115
Figure 4-6	Prediction from a constant PLS model during and after grade transition	116
Figure 4-7	Results from the NIPALS based RPLS algorithm on unscaled data	121
Figure 4-8	Results from the NIPALS based RPLS algorithm on scaled data with constant scaling parameters	122
Figure 4-9	Results from the NIPALS based RPLS algorithm on scaled data with variable scaling parameters	123
Figure 4-10	Results from the kernel based RPLS algorithm using sample	

	wise updating	125
Figure 4-11	Results from the kernel based RPLS algorithm using block wise updating	126
Figure 4-12	Results from the kernel based RPLS algorithm using a moving window length of 10	128
Figure 4-13	Results from the kernel based RPLS algorithm using a moving window length of 5	129
Figure 5-1	Polymerization of ethylene into Polyethylene (top), Structure of Ethylene Vinyl Acetate copolymer obtained from polymerization of ethylene and vinyl acetate (bottom)	133
Figure 5-2	Model Identification for a mixing node	143
Figure 5-3A	Figure 5-3A: Autoclave reactor at AT Plastics (note the scale by comparing the van in front of the reactor)	147
Figure 5-3B	Cross Section of the Reactor	149
Figure 5-3C	Schematic of the Autoclave reactor configuration	150
Figure 5-4	Block diagram to estimate conversion	156
Figure 5-5	Energy balance closure error as estimated from the normal PCA model	159
Figure 5-6	Plot of X during normal operation (used to build the PCA model)	161
Figure 5-7	Plot of scaled X during normal operation	162
Figure 5-8	Figure 5-8: Theoretical value of closure error calculated from mechanistic model (Eq. 5-27)	163
Figure 5-9(a)	1st PC from normal PCA model	165
Figure 5-9(b)	2nd PC from normal PCA model	166
Figure 5-9(c)	3rd PC from normal PCA model	166
Figure 5-9(d)	4th PC from normal PCA model	167
Figure 5-10	Plot of Eigenvalues of the Covariance matrix of X (Scree Plot)	167
Figure 5-11(a)	1st PC from validation data	168
Figure 5-11(b)	2nd PC from validation data	169
Figure 5-11(c)	3rd PC from validation data	169
Figure 5-11(d)	4th PC from validation data	170

Figure 5-12	Error from validation data calculated from theoretical model (Top), and from PCA model (Bottom)	172
Figure 5-13(a)	Squared Prediction Error (SPE) calculated from the PCA model..	173
Figure 5-13(b)	Zoom in of the SPE plot where the decomposition occurs	174
Figure 5-13(c)	Zoom in of PC2 for validation data near the fault	174
Figure 5-14	(a) Energy balance closure error estimated from the PCA model for case study 1; (b) zoomed in to region where fault occurs	176
Figure 5-14(c)	Ethylene Conversion during decomposition (Case study 1)	177
Figure 5-14(d)	Reactor temperature profile during decomposition (Case study 1)	177
Figure 5-14(e)	Initiator Flow rates during decomposition (Case study 1)	178
Figure 5-15	(a) Energy balance closure error estimated from the PCA model for case study 2; (b) zoomed in to region where fault occurs	179
Figure 5-15(c)	Reactor temperature during decomposition (Case study 2)	180
Figure 5-15(d)	Initiator flow rates during decomposition (Case study 2)	180
Figure 5-16	(a) Energy balance closure error estimated from the PCA model for case study 3; (b) zoomed in to region where fault occurs	182
Figure 5-16(c)	Reactor temperature during Case study 3	183
Figure 5-16(d)	Initiator flow rates during Case study 3	183
Figure 5-17	(a) Energy balance closure error estimated from the PCA model for case study 4; (b) zoomed in to region where fault occurs	185
Figure 5-17(c)	Reactor temperature during Case study 4	186
Figure 5-17(d)	Ethylene conversion during Case study 4	186
Figure: 5-18	Combined monitoring index for Case study 1	189
Figure 5-19	Continuous estimation of the closure error as implemented for evaluation at AT Plastics	190

List of Abbreviations

ANN	Artificial neural network
CF	Compression factor
CSTR	Continuous stirred tank reactor
CTA	Chain transfer agent
CUSUM	Cumulative sum
CVA	Canonical variance analysis
DCS	Distributed control system
EVA	Ethylene vinyl acetate
EWMA	Exponentially weighted moving average
FCM	Fuzzy c-means
FTS	Fuzzy Tagaki Sugeno
HDPE	High density polyethylene
HiPP	High impact polypropylene
ICA	Independent component analysis
KCM	Krylov controllability matrix
LCB	Long chain branching
LDPE	Low density polyethylene
LLDPE	Linear low density polyethylene
MDPE	Medium density polyethylene
MI	Melt index
MIMO	Multiple input multiple output
M_n	Number average molecular weight
M_w	Weight average molecular weight
MPC	Model predictive control
MSA	Multi scale analysis
MSE	Mean squared error
MWD	Molecular weight distribution

NIPALS	Non linear iterative partial least squares
NIR	Near infra red
NNPLS	Neural network based partial least squares
ODE	Ordinary differential equations
OLE	Object linking and embedding
OLS	Ordinary least squares
OPC	OLE for process control
PC	Principal component
PCA	Principal component analysis
PCR	Principal component regression
PD	Polydispersity
PE	Polyethylene
PLS	Partial least squares
RBF	Radial basis function
RPEM	Recursive prediction error method
RPLS	Recursive partial least squares
SCB	Short chain branching
SPC	Statistical process control
SVD	Singular value decomposition
UCN	Universal control network

Chapter 1
Introduction

1.1 Introduction

Use of polymers has been growing steadily in many industrial fields, such as automobile, food, apparel, electronics, etc. 1995 world production of plastics was estimated at about 100 million tons (Kiparissides, 1996). In the US, a five fold growth of plastics in two decades (1974-1994) was reported (Rodriguez 1996). With the growth of plastics use in Asia, it is expected that the tremendous growth will continue. This tremendous growth of the polymer industry was attained by exploring new and various plastic applications. For example, polypropylene (PP) is now used for almost all automobiles. Soft drink bottles made from polyethylene terephthalate (PET) have almost completely superseded glass bottles, and polyethylene (PE) plastic bags have replaced the paper bags at grocery stores. Consumers are learning that many plastic products are made from the same polymer. For example, PP used for a core material of instrumentation panel of cars is also used for car batteries, indoor, outdoor carpeting, and polyolefin intimate apparel. PE can be used in producing simple shopping bags and food wrappers to heavy-duty crates and even armor plating for bullet-proof vests. For each use, specific properties of the polymer are needed. In order to meet the demand for such diverse applications of the same base polymer, the polymer industries are producing many different grades of high quality polymers.

In recent years, the pressure from customers for greater grade variation and product diversification has been growing while specification of polymer quality has become increasingly severe. In our competitive world, it is now extremely important for polymer industries to develop technology that can tailor polymer properties and control production plants to maximize product quality as well as production performance and safety. This

thesis deals with some of these issues related to polymer properties and polymer plant safety with emphasis given to high pressure polyethylene processes.

1.2 Polyethylene: Manufacturing Processes and the Product

Polyethylene (PE) is the most widespread and the most studied material in the field of polymers. The world production of PE in 2001 was approximately $50x10^6$ tons. 35% of this production was low density polyethylene (LDPE), produced in high-pressure processes, and 65% was high density polyethylene (HDPE) and linear low density polyethylene (LLDPE) produced in low-pressure reactors (Chemical Economic Handbook, 2002). PE resins are essentially linear polymers with ethylene molecules as the main building block, although most PE molecules also contain branches in their chains. A wide variety of PE resins are produced industrially with differing molar masses, origins and types of branching, and uniformity of branching distribution. However, density and degree of branching are the two most important physical and molecular characteristics of PE and largely determine their end applications. Accordingly, the following grouping is adopted based on density:

LDPE (low density polyethylene): density range from 0.91 to 0.925 g/cm^3
MDPE (medium density polyethylene): density range from 0.926 to 0.94 g/cm^3
HDPE (high density polyethylene): density range from 0.941 to 0.965 g/cm^3

The density of PE is determined by the degree of short chain branching (SCB). The lower the degree of SCB, the higher is the density. Typical branching frequency in LDPE is 10-40 SCB and 0.3-3 LCB (long chain branching) per 1000 backbone carbon atoms, respectively.

Polyethylene is commercially produced by chain growth polymerization using either of the two major processes: high-pressure free radical polymerization, and low pressure coordination polymerization process.

Free radical polymerization is carried out at very high pressure (1000-3500 atm) and high temperature (140-330 ^{0}C) in supercritical ethylene in presence of free radical initiators like peroxides, azo compounds or oxygen. The product is LDPE. Copolymerization with α-olefins and other polar monomers, such as vinyl acetate is also possible. The high pressure process was first discovered in 1933 at the ICI laboratories. Commercial manufacture of LDPE began in the United Kingdom in 1938 and in the United States in 1943.

Coordination polymerization process has been developed more recently (between 1950 and 1976) for the production of HDPE, MDPE and LLDPE (linear low density polyethylene). The reaction is carried out at relatively low pressure (8-80 atm) and temperature (less than 150^{0}C) using a transition metal catalyst (Ziegler-Natta, Phillips or supported metal oxides type). Three commercial processes have been developed for the catalytic polymerization: solution process, liquid slurry (suspension) process, and gas phase process.

PE is probably the most diverse polymer used in numerous fields of applications. Each different type of PE resin has a wide area of application. For example, LDPE produced in high pressure process is a very versatile polymer that is used in a wide range of product applications where its balance of strength, stretch, clarity, sealing, low temperature impact, and adhesion characteristics brings value. HDPE dominates the market of blow molded and injection molded products. LLDPE produced in low pressure process is widely used in film market because of its balance of toughness, puncture resistance, and stiffness. It competes with HDPE and LDPE in many fields of applications because of its superior properties for the applications. Selection of the type of PE resin for any specific application therefore depends on a balance between cost and resin properties. Table 1-1 summarizes the key differences between the high pressure process and the low pressure gas phase process. An industrial study conducted by Nexant Ltd. in 2001 showed that although the investment cost and the energy costs are higher for LDPE compared to the low pressure processes, these costs are offset by the higher comonomer cost required by the low pressure process and higher operating efficiency of the high pressure process.

The cash cost for producing 1 ton of PE is less for the high pressure process. Table 1-2 shows a comparative summary of different types of polyethylene. Excellent reviews on this topic can be found in Kiparissides *et al.* (1993), Ebsworth (2001) and Kissin (2001).

Table 1-1 High pressure process versus low pressure process

	High pressure process	Gas phase process
Reactor residence time	2-3 min (tubular) 20-40 s (autoclave)	1-4 h
Grade change flexibility	Higher	Lower
Monomer cost	Lower	Higher
Energy cost	Higher	Lower
Investment cost	Higher	Lower

Table 1-2: Comparative summary of different Polyethylene resins

	LDPE	HDPE	LLDPE
Process	High pressure (800-1000 atm), high temperature (132-332^{0}C), free radical polymerization	Slurry and gas phase polymerization using catalyst at low pressure (8-80 atm), low temperature (<150^{0}C)	Gas phase, solution or slurry process using Ziegler Natta, Phillips, or metallocene catalyst
Density gm/cm^3	0.915-0.94	0.941-0.96	0.915-0.925
Molecular weight (gm/mol)	about 500 (waxy) to 60000 (tougher products)	about 100 (wax) to several millions (UHMWPE)*	50,000-200,000, narrow mol. weight distribution
Melt temperature (^{0}C)	105 -115	128 -136	120 -130
Optical property	More transparent	Opaque or translucent	Transparent or opaque, depending on branching uniformity
Branching	Both short and long chain branches are present	Linear polymer.	Linear polymer. May contain small number of branches introduced by copolymerization with α -olefins
Use	Largest application in thin film (food packaging, trash can liners etc), also used in extrusion coating. LDPE copolymers dominate the market for adhesives and sealants	Largest use (40%) is in blow-molded products (bottles, toys, drums etc). Next largest (20%) use is injection-molded products (toys, food containers, crates etc). Films such as plastic bag used in food stores, supermarkets etc. represent the third largest market for HDPE	Film is the largest application (grocery, trash bags, non packaging applications in industrial sheeting, agricultural mulch film, and high clarity films used in medical applications and food packaging). Next largest use is injection-molded products for household. Also used for blow molding (toys, large square edged containers such as tanks). LLDPE competes with HDPE in this area of application because LLDPE has much superior environmental stress cracking resistance, and lower gas permeability

*Ultra high molecular weight polyethylene

1.3 High Pressure LDPE/EVA Process Technology

LDPE's discovery was the unexpected result of an experiment undertaken in the course of fundamental research. In the early 1930's without a direct commercial target in view, ICI in England decided to evaluate the effect of ultra high pressure on 50 chemical reactions. An experiment was carried out compressing ethylene gas to 1400 bar. A white solid was formed in the heavy steel vessel, which proved to be LDPE. ICI's Fawcett and Gibson were credited with the discovery and acquired a patent in 1936.

The original reactor was based on an autoclave process using back mixing of the hot reactants with cold incoming ethylene to keep the reactants stable. Later, a tubular reactor process was developed with a plug flow system. In general, polymer produced using tubular reactors result is a narrower molecular weight distribution as compared to the autoclave process. Currently both types of reactor systems are in use commercially. The overall process for the tubular and the autoclave reactor system are very similar except for the design of the reactor itself. The two reactor technologies have the capabilities of producing different products for different end use markets.

A tubular LDPE reactor, which is in essence, a plug flow reactor, consists of a spiral wrapped metallic pipe with a large length to diameter ratio. Total length of the reactor ranges from 500 to 1500 m, and internal diameters can be up to 60 mm. The entire length of the reactor is divided into preheating, reaction and cooling zones. Compressed ethylene gas is injected through the reactor inlet. Additional amounts of the reactant and initiator are fed through the side feed points along the reactor length. The heat of polymerization reaction is removed by an increase in ethylene temperature, by injection of the cold ethylene feed stream, and by heat transfer through the reactor walls to a closed loop water jacket cooling system. Approximately one half of the total heat of reaction is removed through the wall resulting in a non-isothermal reactor. The temperature and flow rate of each coolant stream entering each zone is used to control the temperature profile in the reactor, to maximize ethylene conversion and to optimize product properties. Conversion achieved with this technology ranges from 20 to 30% per pass.

An autoclave reactor is a constantly stirred long vessel with a length to diameter ratio close to 20. The vessel is equipped with an agitator to promote good mixing. The reactor is usually subdivided into different zones. Cold ethylene and initiator solution are fed to the different injection points in the reactor. The multiple zones allow for manipulation of the temperature profile for tailoring of product properties. Reaction conditions can be separately adjusted in each zone to produce polymers with a wide molecular weight distribution. Autoclave reactors behave like an adiabatic continuous stirred tank reactor (CSTR). Addition of the cooler, which is essentially the flow of fresh feed of ethylene and initiator solution, balances the heat of polymerization.

It should be mentioned that most industrial LDPE reactors also produce a number of grades of ethylene vinyl acetate copolymer (EVA). Vinyl acetate has the same reactivity ratio as ethylene and the polymer EVA is produced using the same free radical mechanism as LDPE. EVA copolymers are widely used in elastic films. In addition, owing to their low sealing temperature, high polarity and superior optical property, they are also used in specialty film, foam, adhesive, and extrusion applications. Because of the diverse applications of LDPE and EVA products, the demand for these polymers is steadily growing. Sixty years after its discovery, demand for LDPE produced by the high pressure process continues to grow at 2% annual growth rate (Schuster, 2006).

1.4 Modeling and Control of Polymerization Processes

Molecular and morphological properties of a polymer product strongly influence its physical, chemical, rheological and mechanical properties as well as the end use properties. This makes product quality a much more complicated issue in polymerization processes than in simple chemical systems involving small molecules (Kiparissides 1996, 2004). Polymer resins can be sold only if they meet a set of specifications, which are defined in terms of end use properties such as impact strength, melt index etc. In recent years increasing demand for product diversification in the polymer industry has resulted in more stringent specifications requirements in these properties. Due to the difficulty in measuring many fundamental molecular properties, parameters which are conveniently

measured and thought to be directly related to various end uses are usually reported in the specifications rather than the fundamental molecular properties (e.g. melt index or intrinsic viscosity are reported in lieu of the number average molecular weight, M_n, or weight average molecular weigh, M_w). A major difficulty in applying advanced and automatic control to polymerization reactors arises from the lack of online measurement of polymer properties.

In the polyolefin industry, polymer grades are specified based on values of density and a rheological parameter known as melt index. The melt index value (MI), which has a loose inverse relation to weight average molecular weight, is the weight of a molten resin that flows through a capillary die under a particular stress at a specified temperature (190^0C for polyethylene) for 10 minutes. MI indicates the viscosity of the polymer melt. Online measurement of MI is difficult since it requires close human intervention. Even though online hardware sensors have advanced, the use of new sensors has been largely limited to laboratory reactors (Chien and Penlidis 1990, Kiparissides *et al.* 1998). As a result, in most plants, MI is evaluated off-line and only infrequently, using a manual analytical test. Therefore, the process has to operate without any real time quality indicator between successive measurements. The production of multiple grades of polymer products from the same reactor, common for polymerization reactors, is an additional complication, as it requires frequently changing reactor-operating conditions. Due to the lack of online estimates of product quality, grade changeover typically is a manual operation in most plants, and results in relatively large settling time and/or overshoots. Consequently, a significant amount of off-specification polymer resins are produced. From economic and environmental points of view, it is critical to reduce the amount of off-grade material. This requires sophisticated control systems that are able to provide optimal grade changeover trajectories. The most important factor in developing such a system is the modeling and prediction of quality parameters.

Consequently, a model for estimating polymer properties such as melt index (MI), density etc. using routinely measured process conditions would be extremely useful as an online sensor. Such a relationship, or model, that captures the dynamics between the

process and the product quality variables is often known as a "soft-sensor" since computer software is used to calculate the controlled variable versus using physical instruments. In addition to work as an online estimator, it would allow for optimal scheduling of grade sequence to avoid any mismatch of product quality during grade transitions. Researchers have attempted first principles modeling of polymerization processes to estimate end use properties. However, development of a detailed model derived from first principles, which is capable of continuously predicting melt index from process operating conditions for any polymerization process is not a trivial task. The difficulties involved are the large number of complex and simultaneous reactions and the need to estimate a large number of kinetic parameters. To overcome these difficulties, some researchers considered empirical neural network models as an alternative. However, neural network models provide no physical insight about the underlying process. Data based multivariate statistical regression methods are considered as a viable alternative solution to the problem.

While accurate estimation of polymer properties is essential, ensuring safe and consistent operation of the plant is also of utmost importance in the modern day competitive market. This is particularly crucial for the high pressure polymerization processes that operate under supercritical operating conditions. These reactors require constant monitoring and control in order to minimize any undesirable process excursion that may lead to catastrophic accidents. The production of low-density polyethylene (LDPE) is one such example. During the course of the polymerization reaction, significant amount of heat is liberated as ethylene double bonds are converted to single bonds. The heat of polymerization is roughly 22 kcal/mol, but since molecular weight of ethylene is low, this heat of polymerization is high on a weight basis. This makes the removal of heat of reaction critical on a commercial scale. Under specific conditions outside the normal operating range in the reactor, ethylene can undergo a highly exothermic decomposition reaction and produce carbon, hydrogen and methane. This reaction causes a sharp increase in reactor temperature and pressure and may result in significant economic loss. Therefore, industrial high pressure polymerization reactors are always looking for efficient tools that would be able to detect the occurrence of such unexpected events

ahead of time. There is a significant opportunity for the application of data based multivariate statistical process control strategy to detect and diagnose abnormal shifts in such critical processes. Moreover, in a diverse process plant, where various grades of products are manufactured, these tools are also helpful for product quality control.

1.5 Introduction to Multivariate Statistical Methods

With the development of computer and data storage facilities, nearly every industrial process now routinely collects and stores massive amounts of data on many process variables. Variables such as temperature, pressure, flow rate, etc. are usually measured very frequently, whereas product quality variables such as polymer molecular weight, melt index, density (ρ), long chain branching (LCB), short chain branching (SCB), conversion (x), etc. are measured infrequently. Yet, product quality variables are much more important to the polymer engineer and to the customer. Efficient utilization of the large pool of process variable data can lead to significant improvement in two areas. First, frequently measured process variables can be used to infer the quality variables and an inferential control scheme can be developed. Second, the data can be used to monitor the performance of the process over time for fault detection. In both situations, multivariate statistics play a major role.

Let us consider two sets of measurements X (of size N×n) and Y (of size N×m). X block represents process or causal variables such as temperature, pressure, flow rate measurements etc. Y represents quality variables such as molecular weight, product purity, melt index etc. In many fields of science and engineering, we are often interested to predict the quality variables Y using the routine process variables X with a linear model given by:

$$Y = XC + Noise \qquad 1\text{-}1$$

Such models may be useful in applications such as inferential control. This linear estimator can also be used to model dynamic systems if lagged values of the input and the outputs are included in the X block. In any case, most of the parameter estimation problems that we encounter in engineering practice can be reduced to this simple model form.

The ordinary least squares (OLS) solution of the above system of linear equations is given by

$$C_{OLS} = \left(X^T X\right)^{-1} X^T Y \qquad 1\text{-}2$$

A necessary condition for existence of this solution is that X^TX must be invertible. Mathematically, this is only possible if X is a full column rank, i.e., contains only a set of independent variables. Such an approach creates problems in the presence of correlated process measurements which is often the case with industrial data. In this case, the X matrix becomes rank deficient and in the extreme situation the inverse may not exist. Even if the inverse can be computed, the variance of the estimated parameters will be large indicating that the estimator is unstable. This is known as colinearity or ill conditioning problem of the ordinary least squares method. Moreover, in the OLS solution, it is seen that no attention is paid to the correlation structure of Y. No attempt is taken to reduce the dimension of the X space, which may be required to ensure numerical stability when dealing with redundant measurements. Thus the OLS procedure focuses exclusively on the model fit (predictions) while no consideration is given to the numerical stability aspects of the linear regression problem.

Poor performance of the routinely used OLS procedure in the presence of correlated measurements makes it necessary to look for other available choices. Several multivariate statistical techniques such as PCA (Principal Component Analysis), PCR (Principal Component Regression), and PLS (Partial Least Squares) etc. have been proposed that overcome the shortcomings of OLS. These methods circumvent the collinearity problem by constructing or relating latent or virtual variables which are linear combinations of the original variables. For example, instead of using X and Y directly, we define linear combinations (or latent variables) of the columns of X and Y as follows:

$$t_i = Xj_i, \text{ linear combination of X}$$

$$u_i = Yl_i, \text{ linear combination of Y}$$

The subscript i denotes the i^{th} linear combination of the corresponding space. Further analysis is then performed with these latent or pseudo variables, which are orthogonal and hence uncorrelated to each other. Thus the collinearity problem is overcome.

Moreover, in presence of correlations among variables, usually only the first k ($k<<n$) linear combinations are sufficient to capture most of the variability in the signal space. Therefore incorporating only these few latent variables in the analysis proves sufficient which reduces the dimension of the X space and compresses the data.

The philosophy governing the choice of the latent variables for the X space differentiates these methods. If the goal is to perform data compression and knowledge extraction from only the process variables (X block), then PCA is an appropriate technique. If a numerically stable linear model of the system is sought, then PCR, PLS, CVA (canonical variance analysis) etc are the available choices. In the latter application, two attributes are of major importance for the estimator (i) numerical stability (ii) obtaining good fit of the data. The linear combinations must account for the maximum variation of X and must correlate well with the variables in the Y space to achieve the objectives of model stability and goodness of fit. Each of the multivariate methods accomplishes a different level of balance between these two goals. More details on the theoretical aspects of these methods will be provided in later chapters of this thesis.

In industrial data, or data obtained from any real process, correlations among variables arise from multiple measurements of the same variable using different sensors, or from linear relationships among variables arising from mass or energy balances. When such data is used for building inferential models, ordinary least square regression will usually fail due to colinearity problem as explained earlier. Similarly, when performing the task of process monitoring, application of univariate monitoring tools to analyze such data may produce misleading results. Use of univariate statistical process control (SPC) charts is still quite common in many process industries. Univariate SPC charts such as EWMA (exponentially weighted moving average), CUSUM etc. are used to monitor key process variables in order to detect the occurrence of abnormal situations. By detecting the source of this abnormality, improvements in process operation in terms of safety, waste reduction etc. and consequently product quality can be realized. When such a univariate approach is used to analyze multivariate data, interaction between the variables is not

taken into account. This not only results in misleading process information but also makes the interpretation and diagnosis tasks difficult.

It is in these contexts that multivariate statistical methods such as PCA and PLS are finding increased use in the analysis and archiving of industrial data sets. They have been proven a robust alternative for handling large number of ill-conditioned, highly correlated variables. In essence, these methods can be considered as a data compression technique where the total variance of the process is condensed into a very low dimensional latent subspace. Instead of dealing with the original correlated variables, analysis is performed on the latent variables, which are uncorrelated by construction. This data compression feature provides a low dimensional window into the process and facilitates the tasks of modeling, monitoring and fault detection (Kresta *et al.*, 1991). In addition, when these tools are used for modeling a non-linear process, process non-linearity can be incorporated in the model (if known from process knowledge) by including appropriate non-linear terms in the model or by incorporating non-linear inner relationships for PLS. Consequently, PCA and PLS have found numerous applications in areas of chemometrics, process monitoring, and identification. Some examples include the use of PLS to analyze and monitor an industrial ceramic melter (Wise *et al.*, 1991), and online estimation of quality variables in a desulfurization process (Dayal *et al.*, 1997a, 1997b; Miletic *et al.*, 2004) etc. Numerous articles have been published by researchers from both industry and academia on how PCA can be applied to industrial data for process monitoring (Ricker 1988; MacGregor *et al.*, 1991; Qin and McAvoy 1992; Qin, 1993; Wise and Gallagher, 1996; Kourti and MacGregor, 1998 etc.).

MacGregor and co-workers have done some excellent work with simulated data to show the potential of PCA and PLS in identification and control of polymerization process (Duchesne *et al.*, 2002; Jaeckle & MacGregor, 1998; Skagerberg *et al.*, 1992, etc.). However, in the polymer industry very few industrial applications of PCA and PLS have been published. One of the earliest works on the application of PCA on industrial polymerization process was published by Moteki and Arai (1988). The main focus of their work was to utilize PCA technique towards improving the understanding of the

behavior of a high pressure polymerization process for product quality design. Other examples include the use of PCA and PLS for monitoring and product quality prediction for a batch emulsion polymerization process (Neogi and Schlags, 1998), monitoring a batch polymerization reactor (Kourti *et al.*, 1996), and monitoring of polymerization and spray drying processes (Hergeth Wolf-Dieter *et al.*, 2003). However, few of these papers give process details, details on data processing techniques, and validation of the methods with real data.

1.6 Organization of the Thesis

The main purpose of this research was to investigate the potential of applying data based multivariate statistical methods to develop modeling and monitoring tools for high pressure polymerization processes. We focused on data based methods such as PLS and PCA. This work was done in direct collaboration with a plastics producer (AT Plastics Inc.) that produces a variety of LDPE and EVA products using free radical polymerization. All the analyses were performed using data collected from an industrial high pressure autoclave reactor located at the AT Plastics plant site at Edmonton. Through an extensive literature review and discussion with the plant personnel, three main areas have been identified where potential of the data based statistical methods could be tested:

1. development of data based inferential models to infer polymer quality parameters
2. data based adaptive modeling for time varying processes
3. development of monitoring techniques to detect decomposition for high pressure polymerization reactors

The chapters of this thesis are arranged in accordance with its objectives. A brief literature review on polymer fundamentals is given in chapter 2. Recent developments in the area of polymer property modeling are discussed with emphasis given on the high pressure processes. A number of challenging issues are identified that merit further research.

Chapter 3 presents an industrial application of a PLS-based soft sensor for predicting melt flow index for a polyethylene plant. A brief description of the process and special operating conditions are given. Some difficulties associated with data based modeling of any industrial process are discussed. In particular, issues related to data acquisition, data quality assessment, and model implementation strategies are covered.

Steady state linear or non linear data based models are in many cases sufficient to capture the process behavior within reasonable acceptance limits. However for time varying dynamic systems, the model developed from steady state data may become invalid when the process is running under significantly different operating conditions as compared with the model. As explained earlier, most industrial polymerization reactors produce many grades of products and hence involve grade transition. The process dynamics significantly changes during some of these transitions. In such case, the steady state model needs to be adapted to capture the new input-output correlation structure. Chapter 4 presents a recursive PLS based modeling strategy which is based on krylov controllability based PLS theory (De Ruscio, 2000). Existing recursive modeling techniques are also reviewed and compared with the proposed method. Application results of these methods to model some grade transitions are presented for the specific LDPE reactor.

In addition to ensuring product specifications, it is of utmost importance to ensure safe operation of the process. High pressure LDPE reactors often run into an undesirable situation known as decomposition, which may result in catastrophic consequences. A novel monitoring scheme for the detection of decomposition in high pressure LDPE reactor is proposed in chapter 5. The method combines an overall reactor energy balance and PCA to monitor the energy balance closure error which gives an indication of any impending decomposition in the reactor. Finally a summary of the results obtained during the course of this work and recommendations for future work are presented in Chapter 6.

Chapter 2

Review: Polymer Property Modeling

2.1 Introduction

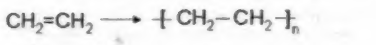
A significant amount of research has been done in the area of control, monitoring, and modeling of polymerization reactors. Results obtained from this collective research have greatly contributed to understand how the reactor operating conditions affect molecular and end use properties. A comprehensive review of all available literature is beyond the scope of this work and only some of the topics pertinent to this work will be covered in this chapter. The focus of this chapter will be on the modeling aspects of the polymerization processes, particularly, on developing inferential systems for polymer properties. This is a very active research area in polymerization reactor control. Excellent reviews on this topic have been done by several researchers (Ray, 1985; MacGregor *et al.*, 1984; Kiparissides, 1996; Ohshima and Tanigaki, 2000 etc.). This chapter starts with some basic but necessary background information on polymer and polymerization processes. The state of the art in quality control systems for polymer production processes is discussed. A detailed review of published literature on modeling polymerization reactors and developing inferential schemes for predicting end use properties of LDPE and related polymers is presented. Since the main focus of this PhD project is high pressure ethylene polymerization processes, emphasis is given on synthesis methods and modeling approaches related to high pressure polyethylene process.

Part of the review material presented in this chapter was published in Sharmin R., Sundararaj, U., Shah, S., Vande Griend, L., Sun, Y., Inferential sensors for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant, *Chemical Engineering Science*, 61, 6372-6384, (2006).

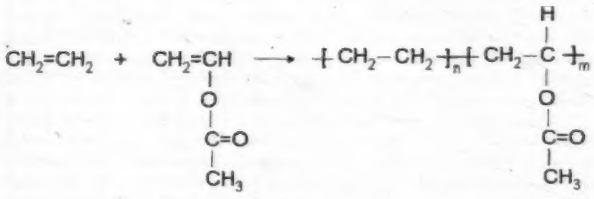
2.2 Polymer and Polymerization Processes

A polymer is a macromolecule made up of large numbers of smaller repeating units. The reaction in which monomers combine to form polymer is termed polymerization. Polymers can be classified according to any of the following criteria: chemical nature of monomers, molecular structure of polymers, polymer chain growth mechanism, or type of polymerization process, polymer texture during use, area of application and so on. Some of these classifications that are relevant to this project are reviewed next (Rudin, 1999).

Homopolymer and copolymer: Homopolymer is a macromolecule derived from a single monomer. Their structure can be represented by multiple repetition of a single type of repeat unit, also known as structural unit. Example includes polyethylene, which is made by linking many molecules of ethylene. A copolymer contains structural units of two or more different monomers. For example, ethylene and vinyl acetate are combined by free radical polymerization process. The product is ethylene vinyl acetate (EVA) copolymer.



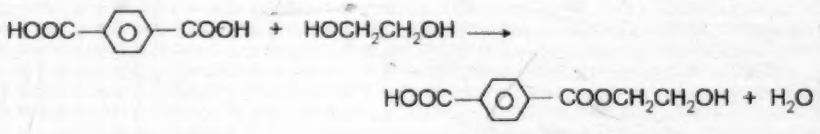
(a)



(b)

Figure 2-1: (a) Polymerization of ethylene into Polyethylene; (b) Structure of Ethylene Vinyl Acetate copolymer obtained from polymerization of ethylene and vinyl acetate

Step growth and chain growth polymerization: Based on the mechanism of polymer chain growth, polymerization reactions can be classified as step growth polymerization and chain growth polymerization. It is possible to place most polymerization processes in one of these two classes, each having distinct characteristics. In step growth process, the growth of polymer molecule proceeds via a stepwise intermolecular reaction. It involves successive reactions between pairs of mutually reactive functional groups which initially are provided by the monomers. Only one type of reaction is involved in the polymerization process. Consider the reaction between terephthalic acid and ethylene glycol, both of which are bifunctional:



The product of this reaction is an ester which possesses one carboxylic acid end group and one hydroxyl end group. This dimer therefore can react with other molecules of terephthalic acid, ethylene glycol or the dimer itself leading to the formation of difunctional trimers or difunctional tetramer. The formation of linear polyesters in this manner is a typical example of step growth polymerization process and is represented in general as follows (here R_1 and R_2 represent any divalent group, usually hydrocarbon):

$$n\text{HOOC–R}_1\text{–COOH} + n\text{HO–R}_2\text{–OH} \rightarrow \text{H[–OOC–R}_1\text{–COO–R}_2\text{–]}_n\text{OH} + (2n\text{-}1)\text{H}_2\text{O}$$

In chain growth polymerization, growth of a polymer molecule is caused by a kinetic chain or reaction (hence the name chain growth). Chain growth polymerization involves reaction of monomers with active centers that may be free radicals, ions, or polymer catalyst bonds. The process is usually initiated by some external source (energy, highly active compounds, or catalysts). The reaction is allowed to proceed under conditions in which monomers cannot react with each other without the intervention of an active center. Chain growth polymerization can be carried out by free radical, ionic or coordination processes. Table 2-1 shows a schematic illustration of the fundamental differences in reaction mechanism between step growth and chain growth polymerization.

Table 2-1: Polymerization mechanism in step and chain polymerizations (modified from Young and Lovell, 2002)

Formation of	Step polymerization	Chain polymerization
Dimer	M + M → M–M	I + M → I–M I–M + M → I–M–M
Trimer	M–M + M → M–M–M	I–M–M + M → I–M–M–M
Tetramer	M–M–M + M → M–M–M–M M–M + M–M → M–M–M–M	I–M–M–M + M → I–M–M–M–M

M: molecule of monomer, I: initiator species

Among the three processes used for chain polymerization, free radical process is the most widely practiced method and is used almost exclusively for the preparation of polymers from monomers of the general structure $CH_2=CR_1R_2$. The process is usually much less sensitive to the effect of adventitious impurities than ionic chain growth reactions. The basic reaction steps involved in the synthesis of polymers by free radical process are briefly summarized next.

Free radicals are independently existing species which possess an unpaired electron and are highly reactive with short lifetimes. In free radical polymerization each polymer molecule grows by addition of monomer to a terminal free radical reactive site known as an active center. The overall reaction can be divided into three distinct stages, namely, initiation, propagation and termination.

Initiation: This stage involves creation of active center and usually takes place in two steps. First, free radicals are introduced into the system. The most common method of generating free radicals involves the use of a thermolabile compound called initiator, which decomposes to yield two free radicals at the reaction temperature.

$$I \rightarrow 2R^{\bullet}$$

Here the initiator I decomposes to yield two radicals $R^{\bullet}$. Organic compounds with bond dissociation energies in the range of 100-165 KJ/mol that contain the O-O peroxide linkage are used as initiators using thermal decomposition method. Other methods of

producing free radicals include redox reactions, photochemical initiation etc. Once free radicals are formed, the initiation reaction follows if a radical R^* adds to a monomer. Initiation reaction in the free radical polymerization of ethylene to form low density polyethylene is represented as:

$$R^* + CH_2{=}CH_2 \rightarrow R{-}CH_2{-}CH_2^*$$

The above reaction is generally written as $R^* + M \rightarrow M_1^*$, where M denotes a monomer and M_1^* denotes a monomer ended radical.

Propagation: This step involves the growth of the polymer chain by rapid sequential addition of the monomer to the active center. There are two major propagation reactions under the conditions of most free radical polymerizations. These are addition and atom transfer reactions.

Addition reactions: successive monomer additions after the initiation step can be represented as

$$M_1^* + M \rightarrow M_2^*$$
$$M_2^* + M \rightarrow M_3^*$$
$$M_i^* + M \rightarrow M_{i+1}^*$$

Here M_i represents the radical $R(\text{–}M\text{–})_{i-1}M^*$. Each reaction in the sequence involves the addition of a monomer to a monomer ended radical. The time required for each monomer addition is of the order of milli seconds. Thus several thousands of additions can take place in a few seconds.

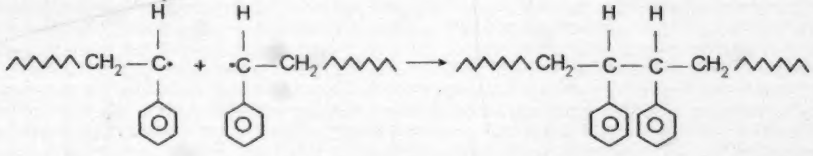
Atom transfer reactions: Radicals can undergo other reactions as well as monomer addition. Atom transfer reaction usually involves transfer of a hydrogen or halogen atom. In free radical polymerization, these reactions are known as chain transfer reactions. Growth of the macroradical is terminated by transfer of an atom to the macroradical from some other species present in the reaction mixture. The donor species itself becomes a radical in the process and the kinetic chain of reaction is not terminated if this new radical can add monomer. In general, transfer of a hydrogen atom between a macroradical M_n^* and a transfer agent TH is written as $M_n^* + TH \rightarrow M_nH + T^*$. The transfer agent TH

can be a monomer, solvent, initiator, polymer or any other substance in the reaction mixture. The new radical $T^{\bullet}$ can reinitiate by combining with a monomer M in the following sequence:

$$T^{\bullet} + M \rightarrow M_1^{\bullet} \rightarrow M_2^{\bullet} \rightarrow \cdots \rightarrow M_n^{\bullet}$$

Chain transfer reactions usually results in a reduction in the number average degree of polymerization ($\bar{x}_n$). Chain transfer to polymer has no effect on $\bar{x}_n$, but it eventually produces branched polymers.

Termination: In this step, the growth of a polymer chain is terminated by mutual annihilation of two radicals. Such termination can occur if two radicals combine to form a paired electron bond as in:



This process is called termination by combination and is represented as $M_n^{\bullet} + M_m^{\bullet} \rightarrow M_{n+m}$. Alternatively two radicals can form two new molecules by a disproportionation reaction in which a hydrogen atom is transferred. Generally, for this case, $M_n^{\bullet} + M_m^{\bullet} \rightarrow M_m + M_n$.

In free radical polymerization, polymer chain microstructure and polymer molecular weight distribution are largely independent of the initiation mechanism and initiator type. One can therefore often predict details of polymer chain microstructure and MWD as a function of reaction variables such as temperature and concentration of species without a detailed knowledge of the initiation mechanism.

Linear and Branched Polymers: Depending on the molecular architecture, polymers are classified as linear and branched polymers. Polymer produced in both step growth and chain growth process may be linear or nonlinear in structure. In a linear polymer, each repeating unit is linked solely in a linear manner to two other units. Examples include

polystyrene, poly(methyl methacrylate), and linear low-density polyethylene (LLDPE). Linear polymers may contain short branches, which are part of the monomer structure such as those in LLDPE. Branched polymers are those in which the polymer contains polymerized branches. Branching occurs either because at least one of the monomers have functionality higher than two or polymerization process itself produces additional branching points in a polymer made with bifunctional monomers. A major example of the second type of branched polymer is polyethylene that is made from free radical polymerization at temperature between 100~300^{0}C, and pressure of 1000~3000 atm. Depending on the reaction conditions, these polymers contain 20 to 30 short chain branches (ethyl and butyl branches) per 1000 carbon atoms and one or a few much longer branches per molecule. They differ sufficiently in properties from linear polyethylene such that the two materials are generally not used for the same applications.

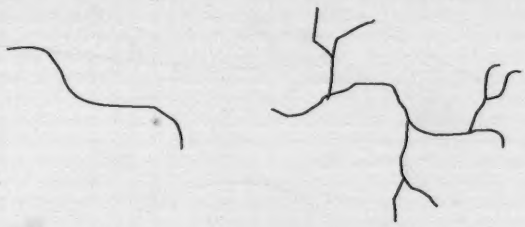


Figure 2-2: Linear and Branched Polymer

2.3 Quality Control in Polymer Industry

Molecular and morphological properties of a polymer product strongly influence its physical, chemical, rheological and mechanical properties as well as end use properties. This makes product quality a much more complicated issue in polymerization process than in conventional reactions involving short molecules. Figure 2-3 and Table 2-2 describe some qualitative relationships between molecular properties, mechanical and chemical properties, and processability of the polymer (Ohshima and Tanigaki, 2000). It is evident that end use properties depend strongly on the polymer low-order and high order structure as well as their distributed nature of variables (molecular weight distribution etc.). For example, hardness of polymer depends on polymer crystallinity that

is determined by stereoregularity of polymer. A linear polymer with small side groups is highly crystalline. The presence of bulky but regular side groups in a linear polymer results in low crystallinity. Bulky and random side group present in a linear polymer makes them non-crystalline. Similarly, the morphological form of a HiPP polymer is often a key variable of end use properties, and it depends on particle size distribution, polymer composition distribution, and processing history in the extruder. Polymers differ from small sized compounds in that they are polydisperse or heterogeneous in molecular weight. MWD of a polymer is a record of the kinetic history of the reactions which occurred during its formation (Clay and Gilbert, 1995). Therefore it is an important source of information about the kinetic processes that have taken place in polymerization systems. Control of molecular weight and MWD is often used to improve certain desired mechanical and physical properties of a polymer product.

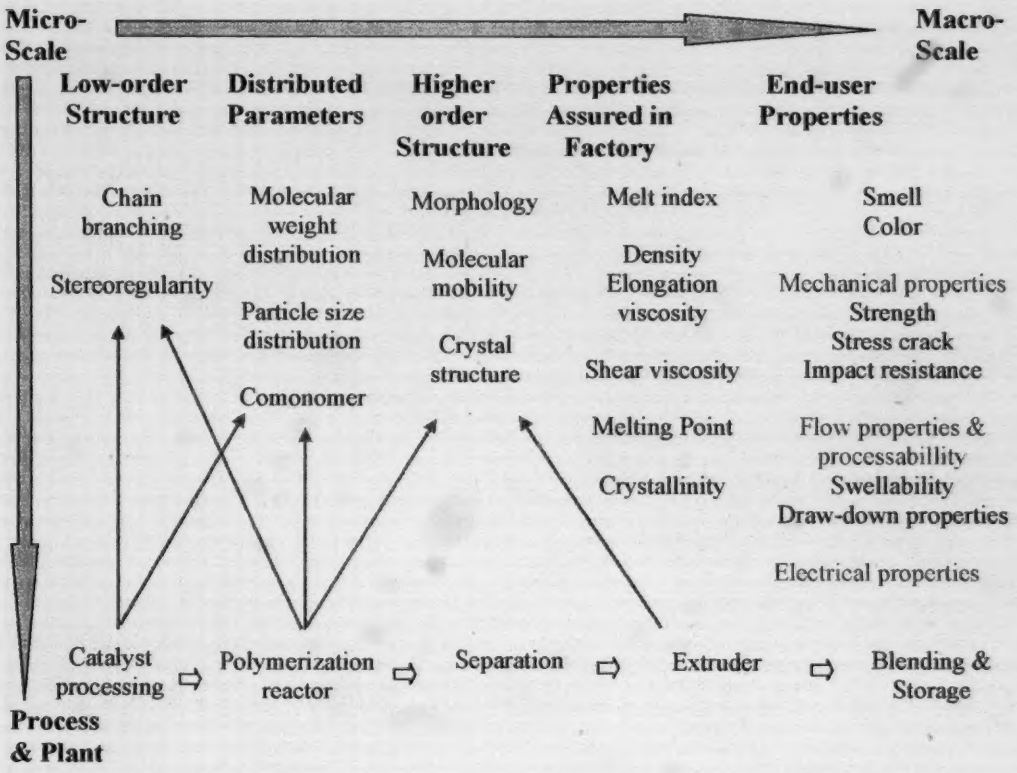


Figure 2-3 Relationships among polymer properties and processing condition (adapted from Ohshima and Tanigaki, 2000)

Table 2-2: Relationship between molecular and physical properties of polymer (adapted from Ohshima and Tanigaki, 2000)

Molecular structure ⟶		Molecular weight (M_w)	Molecular weight distribution	Branching chemicals	Degree of branching	Degree of branching distribution	Long chain branching
Mechanical and chemical properties	Transparency	x	x	x	x	x	
	Tensile strength	x	x	x	x	x	x
	Impact strength	x	x	x	x	x	x
	Rigidity/Flexibility				x	x	
	Heat resistance				x	x	
	Cold resistance	x	x	x	x	x	
	Chemical resistance	x	x	x	x	x	x
	Heat seal	x	x	x	x	x	
Processability	Bubble stability	x	x				x
	Draw down	x	x				x
	Extrusion torque	x	x				x

In industrial settings, polymer resins typically need to be approved using a quality control protocol. Resins can be sold only if they meet a set of specifications which are defined in terms of end use properties such as impact strength, melt index, density etc. Due to the difficulty in measuring many fundamental molecular properties, those parameters, which are conveniently measured and thought to be directly related to various end uses, are usually reported rather than fundamental molecular properties (eg. MI or intrinsic viscosity is reported in lieu of number average molecular weight M_n, or weight average molecular weight M_w). However, online measurement of these end use properties is time consuming and requires close human intervention. Some advancement was made in developing on-line sensors to measure polymer properties at the extruder units. On-line sensors for rheological properties such as viscosity and yield stress were developed (Broadhead *et al.*, 1993). Using fiber optic linked devices, such as Raman and NIR, new sensors were also developed for monitoring polymer properties (Kettry and Hansen, 1995; Chien and Penlidis, 1990; Kiparissides *et al.*, 1998). Among them, near infrared (NIR) spectroscopy has greatly expanded capabilities with the help of recent advances in chemometrics. Watari *et al.*, (1996) utilized the NIR to measure the density of polyethylene on-line. However, use of these online sensors is very limited and until now, in most polymer plants, these properties are usually measured offline, and are available infrequently.

Lack of online estimates of key quality variables makes the quality control of polymer plants difficult. Consequently, a model for estimating polymer quality parameters using current process conditions would be very useful for online inferences. Moreover, it is indispensable to have deeper understanding about how the end use properties are affected by molecular structure, molecular weight distribution and operating conditions of each processing units, in order to determine the operating conditions for each individual processing unit. The quantitative representation of the relationship is called the quality model. To construct a quality control system for a polymerization plant it is crucial to develop such quality models which can describe the relationship among end user properties and polymer properties, and molecular structures and processing history.

Recent literature on this topic suggests that development of inferential system for polymer properties is a very active research area in polymerization reactor control (Chan *et al.*, 1993; McAuley, 1990; Kiparissides *et al.*, 1993; Zabisky *et al.*, 1992; Chan and Nascimento, 1994; Skagerberg *et al.*, 1992 etc). The models used to infer polymer properties can be roughly categorized into three groups: (1) mechanistic models developed from first principles, (2) black box models using neural networks, and (3) statistical models using multivariate statistical tools. A brief review of research work published in these three areas is presented next.

2.3.1 First principles models

A major objective of the polymer reactor modeling is to understand how the reaction mechanism, physical transport phenomena, reactor type and reactor operating conditions affect the polymer quality of the final product. A large number of computer models have been developed for high-pressure LDPE reactors over the past two decades. Table 2-3 summarizes some of the major publications on LDPE reactor modeling. These models differ in various degrees of complexity but the fundamental approach taken by different researchers is very similar. First, the entire reactor length is divided into a large number of small discrete segments. Then, a set of elementary reactions is chosen to describe the polymerization process for each segment. These include initiation, propagation, and termination steps. Next, the differential equations describing (a) mass balances of various "living" and "dead" polymer chains, monomer, initiator, solvents, (b) energy balances for the reaction mixture and the cooling (or heating) fluid, (c) molecular properties (long chain branching, short chain branching, comonomer content etc.) and (d) velocity along the reactor are formulated and solved simultaneously. The number of molar balance equations that need to be solved depends on the degree of polymerization, and the total number of equations can be as high as 10^4. Several mathematical techniques have been proposed in order to reduce the infinite set of differential equations to a lower order system which can be solved numerically (Ray, 1972; Mills, 1986, etc.) out of which, method of moments (Hulburt and Katz, 1964) is the most widely applied. Solution of these coupled non-linear ODE's gives fractional conversion of various species, temperature profile, moments of chain length distribution of living and dead chains along

the reactor length. Fundamental molecular properties such as number and weight average molecular weight, long and short chain branching, and comonomer composition can be derived from the polymer moments. Some of the modeling works summarized in Table 2-3 are discussed in more detail next.

Using this approach, Kiparissides *et al.*, (1993) developed a comprehensive reactor model for a LDPE tubular reactor and derived temperature profile, conversion, and polymer properties such as M_n, M_w, SCB, and LCB along the reactor length. To estimate the melt index and density, the following empirical relationships were incorporated in the model:

$$\log MI = a + b\,LCB + c\,\log M_w ; \qquad Density = \alpha + \beta SCB$$

Through extensive simulation, their model examined the effect of varying initiator efficiency and fouling factor on reactor operation and properties. The simulated temperature profiles at different times were compared with actual measurements.

Chan *et al.*, (1993) also developed a mathematical model to study free radical polymerization and copolymerization in high-pressure autoclave reactors producing LDPE and EVA. They used actual industrial recipes to simulate the temperature profile, comonomer composition, total initiator flowrate, and % conversion for two grades of LDPE and EVA resins produced in industry. Their result showed good agreement with the properties measured for the industrially produced resins. However, it should be noted that comparisons are only made on an average cumulative value of these properties. The model parameters were tuned separately to match the data for the two types of resins.

Table 2-3: Summary of literature on high pressure LDPE reactor modeling

References	Reactor Type	Summary
Feucth *et al.*, (1985)	Vessel	Detailed mathematical model on autoclave reactors. Prediction of molecular properties of LDPE
Marini & Georgakis (1984a,b)	Vessel	Investigation of mixing phenomena in LDPE reactors. Prediction of initiator productivity and polymer quality.
Tilger and Luft (1988)	Tubular	Two-dimensional dynamic model developed for high pressure LDPE reactor. Variation of physical properties along the reaction coordinate is considered.
Zabisky *et al.*, (1992).	Tubular	A copolymerization model for tubular reactor is proposed. Model used to simulate operation of commercial reactor.
Kiparissides *et al.*, (1993)	Tubular	A comprehensive mathematical model is developed for homopolymerization of ethylene in a 2-zone tubular reactor with intermediate feed
Verros *et al.*, (1993)	Tubular	A mathematical model based on double moments is employed to calculate the molecular weight and compositional changes for copolymerization of ethylene in a 2-zone tubular LDPE reactor
Chan *et al.*, (1993)	Vessel	A copolymerization model for vessel reactor is developed. Two phase flow and gel formation from cross linking reactions are taken into account. Model is used to simulate the operation of commercial reactors.
Ham and Rhee, 1996	Vessel	A two-compartment four-cell simulation model is developed to investigate reactor performance represented by monomer conversion and reactor temperature. Polymer properties such as M_w and polydispersity (PD) are also predicted.
Pladis and Kiparissides, 1998	Vessel	A comprehensive mathematical model is developed to calculate joint molecular weight-long chain branching distribution of highly branched polymers produced in free radical process. Effect of polymerization temperature, CTA concentration, and reactor residence time distribution on MWD of LDPE is investigated.
Zhou *et al.*, 2001	Tubular/ Vessel	Detail mathematical model for LDPE reactor is developed using method of moments and computational fluid dynamics. Prediction of monomer conversions, polydispersity, radical distribution, and MWD. Influence of initiator concentration and inlet temperature on conversion and PD is investigated.
Alleyne, 2006	Vessel	A comprehensive first principles model is developed to predict product qualities for steady state and grade transition. Effect of imperfect mixing, and gell effect was investigated.

Computer models provide a sound basis to mathematically describe the high-pressure LDPE reactors. These models give excellent insight into how reactor configuration, reaction kinetics and reactants affect various operating conditions and polymer properties. The following remarks hold for implementing a model to predict polymer quality:

1. The molecular balance equations require the reaction rate constants for all the elementary reactions. In addition, the effect of the reactor operating conditions on physical, thermodynamic and transport properties of the reaction mixture must be known. For any comprehensive modeling study, the resulting parameter set that needs to be evaluated either from experiment, or from established literature, becomes too large to manage.

2. Given a set of reactor operating conditions, continuous prediction of properties related to the end use of polymer such as MI, is not straight forward using these models. No explicit relationship between the end use properties and the operating conditions can be obtained. It is not clear how these models can be applied online for real time quality estimation.

McAuley and MacGregor (1991) developed a theoretically based model for the production of HDPE in a fluidized bed reactor to relate M_w and comonomer composition using online temperature and gas composition measurements. An empirical relationship between M_w and MI was incorporated into this model:

$$\ln(MI) = k_7\left(\frac{1}{T} - \frac{1}{T_0}\right) + 3.5\ln\left(k_6 + k_2\frac{[M_2]}{[M_1]} + k_2\frac{[M_3]}{[M_1]} + k_3\frac{[H_2]}{[M_1]} + k_4\frac{[R]}{[M_1]} + k_5\frac{[I]}{[M_1]}\right)$$

Here $[M_1]$, $[M_2]$, $[M_3]$, $[H_2]$, [I], and [R] are concentration of monomer, comonomers, hydrogen, cocatalyst, and impurities in the gas phase in the reactor. Parameters k_1 to k_4 and k_7 were estimated using steady state offline data. The remaining model parameters were updated using offline MI measurements by recursive prediction error method (RPEM). The model was able to capture melt index and density of HDPE resins during both steady state operation and grade transitions. This model was later used (sometime

with minor modifications) by many researchers in their study of HDPE and LLDPE reactors (Ogawa *et al.*, 1999; Ohshima *et al.*, 1995; Ohshima and Tanigaki, 2000; Sato *et al.*, 2000). Direct application of this model for LDPE production is not suitable for two reasons:

1. HDPE is produced using a coordination polymerization process where copolymerization of ethylene and α-olefins are carried out in a fluidized bed reactor at low pressure using a heterogeneous Ziegler-Natta or supported metal oxide catalyst. LDPE Production is performed in a free radical high-pressure polymerization process using initiators. The basic reaction mechanisms are quite different.

2. One major assumption used in McAuley's model is the uniform mixing of the reactants in the gas phase of the reactor, which allows them to use a uniform temperature and gas composition throughout the reactor. Georgakis and Marini (1982) studied the mixing pattern in LDPE autoclave reactors. The initiator tends to decompose near the feed points and not in the bulk of the reactor; therefore, less polymerization occurs compared to what would have been achieved if there were uniform distribution of initiators throughout the reactor. This results in non-uniform composition and develops a temperature gradient along the reactor. Another practical issue that is worth mentioning is that the gas composition inside the reactors usually cannot be measured for high-pressure LDPE reactors. Essentially, it is not feasible to obtain gas samples from a reactor that is operating at pressures above 1200 atm.

In a recent work Alleyne (2006) developed a comprehensive first principles model for high pressure stirred autoclave reactor in a LDPE/EVA polymerization plant. The model included unit process based equation of state thermodynamics and full free radical kinetics for all reacting components and was used to predict reactor temperatures, conversion, final product qualities such as comonomer content and melt index. Theoretical additions made by other researchers such as Trommsdorff (gel) effect, and

effect of imperfect mixing were incorporated into this model. Validation results were shown for the steady state reactor, steady state full plant model, and dynamic full plant model. The steady state model was validated with actual plant data for a number of grades of product. The dynamic model was validated using data from a typical grade transition in the plant. The model showed an excellent fit with experimental data in all cases. Finally, the model was used to develop optimal grade transition policies which were implemented in the actual plant. Although a great majority of the reactor models published in the literature were built using computer programs such as FORTRAN, MATLAB, or SIMULINK, this model was built using a commercial software package (Aspen Polymer plus) in close collaboration with the plant personnel to ensure that it was practical, usable and sustainable. This is an excellent contribution towards modeling of high pressure polyethylene reactors.

2.3.2 Black Box Models: Use of Neural Networks

There is a significant potential for employing neural networks in chemical engineering applications since all major chemical processes are highly non-linear. Neural networks have the ability to approximate any continuous non-linear functions and have been applied to non-linear process monitoring (Bhat and McAvoy, 1990; Qin and McAvoy, 1992; Chan and Nascimento, 1994; Zhang *et al.*, 1998; Rallo *et al.*, 2002 etc.). They possess the ability to learn what happens in a process without actually modeling the physical and chemical laws that govern the system. The success in obtaining a reliable and robust network depends strongly on the available set of data, the choice of process variables and its training domain (Bhatriya and Whiteley, 2001). Table 2-4 summarizes some of the major publications on application of neural networks in polymerization processes.

Table 2-4: Summary of literature on Neural Network based modeling of polymerization processes

Reference	System	Summary
Chan and Nascimento, 1994	LDPE tubular reactor, industrial study	Neural network is used to model melt index, density, Mn, Mw, and conversion
Ohshima *et al.*, 1995	HDPE reactor, industrial case study	Melt index is predicted using artificial neural net (L^{∞} wave net). The model by McAuley and MacGregor (1991) is used for input selection
Goiochen *et al.*, 1995	Batch MMA reactor, laboratory scale study	Polymer chain length is estimated from concentration and conversion data using extended Kalman filter and neural network.
Zhang *et al.*, 1998; Zhang, 1999	Batch MMA[1] reactor, simulation study	Bootstrap aggregated neural network is used to estimate M_n and M_w
Ogawa *et al.*, 1999	HDPE slurry reactor, industrial case study	MI estimated using L^{∞} wave net (same model as in Ohshima et al, 1995). The inferential system is combined with a quality control system that utilizes a two degrees of freedom cascaded MPC[2].
Rallo *et al.*, 2002	LDPE tubular reactor, industrial case study	Single grade and composite neural network models are built using fuzzy ARTMAP neural system. Details on data preprocessing is given.

[1] MMA: methyl methacrylate

[2] MPC: Model predictive controller

In an application related to polymerization reactors, Chan and Nascimento (1994) used a back propagation neural network to model free radical polymerization in a high pressure LDPE tubular reactor. They used the reactor pressure, mass flow rate of monomers and initiator feeds, and feed temperature to model temperature profile along the reactor, melt index, density, M_n, M_w and monomer conversion. Industrial data was used to train the network. The results show good agreement for all the properties for one single grade. It should be mentioned that the network was not applied for continuous estimation of any of these properties. Instead, they compared a cumulative average value for one LDPE resin. In a more recent work, Rallo *et al.*, (2002) employed a predictive fuzzy ARTMAP neural system to estimate MI for six different LDPE grades produced in a tubular reactor. Sensitivity analysis with self-organizing maps was used to select the most relevant process features and to reduce the number of input variables to the model. They developed separate models for each grade and a composite model for all six grades. These models could successfully capture the time trend for polymer properties at steady state and during grade transitions.

The above discussion shows that neural networks appear to be quite an attractive method to infer polymer quality, but in practice, only a very limited number of industrial applications have been reported. One disadvantage in using neural networks for data mining is its slow learning process. The technique usually requires a large sample size and trial and error. Therefore, training can be quite time consuming. Another disadvantage is that neural networks do not give explicit knowledge representation in the form of rules, or some other easily interpretable output. The model is implicit, and knowledge is hidden in the network structure and optimized weights between the nodes. The individual relationships between the input and the output variables are not developed by using any engineering judgment, therefore the model tends to be like a black box. That is, the input-output table has no analytical basis. This is the major reason why there is very limited acceptability of this method in industrial practice, particularly in polymerization processes. It may be possible to train a neural net using data generated from a valid mechanistic model. This approach could be beneficial in applications such as

product design using model inversion techniques. However, to the best of author's knowledge, no such application has been published.

2.3.3 Data based Statistical Models

Multivariate statistical methods such as PCA, PCR and PLS have been successfully applied to solve a wide range of multivariate problems in chemometrics, and have recently gained importance in chemical engineering. An important aspect of these robust regression techniques is their ability to handle large number of ill-conditioned, highly correlated variables. Ordinary least square regression will usually fail in these cases. Consequently, these methods have gained rapid acceptance and utilization in industry. Examples include the use of PLS to analyze and monitor an industrial ceramic melter (Wise *et al.*, 1991), monitoring a batch polymerization reactor (Kurti *et al.*, 1996), and online estimation of quality variables in a desulfurization process (Dayal and MacGregor, 1997a, 1997b; Miletic *et al.*, 2004).

In polymerization processes, the use of multivariate statistical methods to develop inferential models is limited to mainly simulation studies. A series of articles has been published by MacGregor & co-workers and Kiparissides & co-workers in this area. Skagerberg *et al.*, (1992) and MacGregor *et al.*, (1994) demonstrated the use of PLS in a simulation study, to model M_w, M_n, LCB, SCB, and concentration of vinyl and vinylidene groups for a single grade of LDPE produced in a tubular reactor using the reactor temperature profile and the solvent flow rates as the input. Although their model was shown to perform well, it was only applicable to a single grade. The results were not validated with any real plant data. In addition, the issue of multiple grades and grade changes, which are very common to polymerization reactors, was not covered in their work.

Since many chemical processes exhibit nonlinear behavior, a number of non linear PLS algorithms have been developed to capture the true nonlinear dynamics. These algorithms use different nonlinear approaches to model the inner relationship between each pair of input- output latent variables. For example, Wold *et al.*, (1989) proposed a quadratic

function for mapping the inner relationships in PLS. In a more recent study, Jaeckle and MacGregor (1998) used a PLS based method to find a window of operating conditions to obtain a product with a desired set of quality specifications. As a starting point of their methodology, which is known as inverse PLS modeling, they developed a non linear PLS model for M_w, M_n, LCB, SCB and conversion for the production of LDPE in a tubular reactor. The inputs for their model included reactor pressure, feed temperature, initiator and solvent flow rates. Nonlinear inner relationships were used to model the input-output latent variables pair, although the nature of the nonlinearity was not explicitly mentioned in their work. Martin *et al.*, (1999) reported a similar application of inverse PLS to a pilot scale batch methyl methacrylate (MMA) suspension polymerization reactor. In both cases (Jaeckle and MacGregor, 1998; Martin *et al.*, 1999), the entire data were generated using a simulation model developed by Kiparissides *et al.*, (1993) and no validation with industrial data was done. Jaeckle and MacGregor's (1998) model was developed using data combined from nine LDPE grades from the same family of products (e.g. film grade polyethylene). Unfortunately they give little detail about the pre-processing steps, which are crucial when dealing with non-stationary data. In a recent study Jaeckle and MacGregor (2000) applied the inverse PLS method on historical data taken from two industrial polymerization processes. Although the method was not implemented, the results could serve as a starting point for further experimentation.

In addition to the three basic modeling strategies explained so far, another more recent trend in polymer property modeling is noticed where the statistical methods are combined with machine learning methods such as artificial neural networks (ANN). Data based methods such as PCA and PLS are more suitable for modeling linear stationary processes whereas neural network has found wide applications in process identification due to its excellent ability in representing arbitrary nonlinear relationships. Therefore, it is expected that one method will complement the other when used in combination. In support of this proposition, Qin and McAvoy (1992) proposed neural network based PLS (NNPLS) where a neural net is used to model the nonlinear inner relationships. Baffi *et al.*, (1999a, b) modified the input weights updating procedure of the quadratic PLS proposed by Wold *et al.*, (1989) and neural network based PLS algorithms proposed by Qin and McAvoy

(1992). These modifications were proven to improve the modeling capabilities over the original nonlinear PLS algorithms as evidenced from some interesting applications in polymerization processes.

In a recent work by Shi *et al.*, (2006), a novel soft sensor architecture was proposed for an industrial propylene polymerization process. The method combines artificial neural network (ANN) with independent component analysis (ICA) and a multi-scale analysis (MSA) to infer melt index of polypropylene from available process variables. ICA is a general purpose projection technique in which the observed random data, usually containing highly correlated variables, are linearly transformed into components that are maximally independent from each other (Hyvarinen and Oja, 2000; Hyvarinen, 2002). In this study, ICA was used to select relevant features needed as input to the model. In addition, it helped in reducing the dimension of the input space, which simplified the neural architecture and reduced the time required for training. Next, multi-scale analysis (MSA) was applied to obtain more information from the data and reduce the system uncertainties. MSA uses wavelets and corresponding scaling functions basis functions to decompose a set of data into components described by wavelet coefficients. Signal decomposition with MSA has near optimal properties in a wide range of non-homogenous function spaces (Baseville *et al.*, 1992; Li *et al.*, 2004). Finally radial basis function (RBF), which is a typical feed forward type artificial neural network (ANN), was used to model the nonlinear relationship between MI and the process inputs. The results showed acceptable prediction capability for both training and test data sets. The results also suggested that the use of ICA in combination of the neural network (ICA-MSA-RBF) models results in improved predictive performance as compared to the neural network model developed with the original variables. The paper provides sufficient details about the data collection and preprocessing techniques such as scaling, duration of training data etc. which are often ignored in the literature involving real plant data. The method seems to have promising potential for practical use. However, the results presented suggest that the analysis was performed with a single grade of product. Hence issues of multiple grades, grade transitions, and practical considerations during implementation etc were not discussed.

However, the neural network based nonlinear PLS methods suffer from criticisms such as over fitting the data and lack of model interpretability. In order to improve model interpretability, Bang *et al.*, (2003) proposed to incorporate the fuzzy Takagi-Sugeno (FTS) model for the inner mapping, named as fuzzy PLS. In a more recent work Liu (in press) developed a soft sensor algorithm for a nonlinear time varying process by combining PCA, fuzzy c-means clustering, and FTS modeling. The underlying principle of this method is that a nonlinear system can be decomposed into several local subsystems using fuzzy c-means clustering method meanwhile building a linear model for each local subset assuming that the local input/output relationship is linear. PCA was used to eliminate collinearity and reduce the input variable dimensions. Next, FCM was used to decompose the operating space into several local regions. And finally FTS model is used to build a local linear model for each region. The method was successfully applied to predict melt index data from a high pressure polymerization process producing three different grades of polymer product.

2.4 Summary: Challenges & Motivation

Development of detailed first principles model for free radical polymerization is not a trivial task. The difficulties involved are the large number of complex and simultaneous reactions, poor understanding of flow patterns in the reactor and poor understanding of heat and mass transfer of the polymer monomer mixtures, and the need to estimate a large number of kinetic parameters. Although a great number of papers have been published on the modeling of LDPE reactors, a consistent set of kinetic rate constants has not been established in the open literature. This may be attributed to the complexity of the reaction mechanism, the large number of kinetic parameters to be identified experimentally, and the wide range of experimental conditions over which the kinetic parameters are estimated. The dependence of thermodynamic and transport properties of the reaction mixture (i.e., density, specific heat, viscosity, thermal conductivity) on pressure, temperature and composition must be known in any comprehensive modeling study. In light of these difficulties many researchers consider empirical neural network models as an easier alternative. However, it provides no physical insight about the underlying

process. Therefore, due to limited acceptability of this approach, in this work, we will focus on data based statistical methods.

Although there is great potential for applying PLS to inferential modeling, it is evident from the available literature that very little work has been done in the area of polymerization. To the best knowledge of the author, no industrial application of PLS has been published to infer polymer quality parameters. However, a number of challenging issues are involved in direct application of PLS to model polymerization processes. A major complexity arises from frequent grade changes required to produce a large number of different products. Even for a single reactor system, operating conditions and polymer properties can vary over a wide range. Polymer quality parameters such as MI are believed to have non-linear relationships with operating conditions, but the explicit nature of non-linearity is still unknown. Bremner and Rudin (1990) developed an empirical relationship to relate MI with M_w for LLDPE and HDPE. Their model had some validity for linear polymers of similar molecular polydispersivity and processing history. For branched polymers (LDPE), they concluded that the situation is more complex, and no general relationship can be obtained unless the rheology of the polymer can be characterized. The process dynamics also change during grade transition. Sato *et al.*, (2000) examined grade transition through a simulation study and showed that the process gain can change greatly due to product change. These issues suggest that a single PLS model will be insufficient to capture the true process dynamics unless it is developed within an adaptive framework.

It should also be pointed out that there are some inherent difficulties in data based models using closed loop industrial data as opposed to the data generated from simulation. The data acquisition and preprocessing, and proper variable selection can be very demanding and time consuming. Process knowledge has to be intelligently combined in this step for model success. There is a significant lack of literature on data based statistical models for polymerization processes. A limited number of articles concerned with grade transition do not clearly describe the data preprocessing techniques, which is vital for the ultimate success of the model. For example, the issue of scaling in this case is still an unresolved

problem. Another disturbing aspect of the previous work is the tendency to evaluate the model performance based on the training data set rather than applying the model to a new data set that is unseen by the model (Chan and Nascimento, 1994; Qin, 1998). In terms of research needs, much work remains to be done before a generalized framework can be accepted for data based modeling of time varying non-linear processes.

Chapter 3

Data Based Inferential Modeling

3.1 Introduction

Multivariate statistical tools such as PLS and PCA have been used to analyze data for process monitoring and developing inferential models in a variety of disciplines such as science, social science, engineering and medicine. Although application of these methods in the field of polymer is not too wide spread, it is interesting to note that perhaps the first application of the PCA technique in an industrial setting was reported by Moteki and Arai (1986), who used it to derive optimal operating conditions to synthesize specific polymer grades. Since then many researchers have demonstrated through numerous simulation studies that there is great potential for application of latent variable based regression techniques to predict polymer quality parameters using available process data. However, process nonlinearity, poorly known nature of relationship between final polymer quality with reaction conditions, and the inherent difficulties associated with dealing with industrial data pose some challenge to the direct application of these techniques. To the best knowledge of the author, no industrial application of PLS has been published to model polymer quality parameters. In this work, PLS was used to build a soft sensor to predict melt flow index using routinely measured process variables.

Some of the results in this chapter was published in Sharmin R., Sundararaj, U., Shah, S., Vande Griend, L., Sun, Y., Inferential sensors for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant, *Chemical Engineering Science*, 61, 6372-6384, (2006).

The study was conducted using data collected from an industrial autoclave reactor, which produces low-density polyethylene (LDPE) and ethylene vinyl acetate (EVA) copolymer using free radical polymerization. This chapter provides a detailed description of the various steps involved in the soft sensor development starting from data collection to the final implementation of the model. A brief overview of the PLS theory is presented. Some difficulties associated with the analysis of industrial data are also discussed.

3.2 Partial Least Squares: Theory

Partial least squares or projection to latent structures (PLS) technique is a robust alternative to the ordinary least squares (multiple linear regression) method in the analysis of correlated data. This is a latent variable based method used for the linear modeling of the relationship between a set of response variables Y (of size N×k) and a set of predictor variables X (of size N×n). The linear model can be represented as follows:

$$Y = XC + E \tag{3-1}$$

In PLS, the objective is to arrive at a stable estimate of C while performing data compression on both the X and Y blocks. The correlation structure between the X and Y blocks is considered in the estimation procedure. The latent variables for the X block are constructed with reference to the Y space. Therefore a compromise solution is reached that takes into consideration both the stability (through constructing uncorrelated latent variables) and the model fit aspects (through the construction of the X block latent variables with reference to the Y space) of the regression problem (Lakshminarayanan 1997).

The ordinary least squares solution of C is given by

$$C = \left(X^T X\right)^{-1} X^T Y \tag{3-2}$$

A common problem with this solution is that X^TX can become ill-conditioned and the inverse may not exist when X contains many correlated variables. PLS regression is used to overcome this problem (Lakshminarayanan *et al.*, 1997; Di Ruscio, 2000) in the following manner:

We can define

$$t_1 = Xw_1 \tag{3-3}$$

where t_1 is the latent variable in X-space and

$$u_1 = Yl_1 \tag{3-4}$$

where u_1 is the latent variable in Y-space. t_1 and u_1 are also known as the scores vectors in X and Y space respectively. w_1 and l_1 are weight vectors of unit length that will maximize the covariance between t_1 and u_1. The estimation of the weight vectors can be considered as a solution of the following constrained optimization problem (Hoskuldsson, 1988):

Objective: Find a linear combination of X and a linear combination of Y that have the maximum covariance between them.

Objective Function: $\max\{w_1^T \Sigma_{XY} l_1\}$

Constraints:

$$w_1^T w_1 = 1$$
$$l_1^T l_1 = 1$$

Solution: w_1 is the first left singular vector and l_1 is the first right singular vector of X^TY.

In view of the above formulation, the PLS procedure can be analyzed as an eigenvalue eigenvector problem. The weight vectors for each PLS dimension can be obtained from a singular value decomposition (see Appendix 1) of the covariance matrix X^TY. In fact, Hoskuldsson (1988) showed mathematical proof that PLS is an eigenvalue-eigenvector problem where *u, l, t,* and *w* are eigenvectors corresponding to the largest eigenvalue of YY^TXX^T, Y^TXX^TY, XX^TYY^T, and X^TYY^TX respectively.

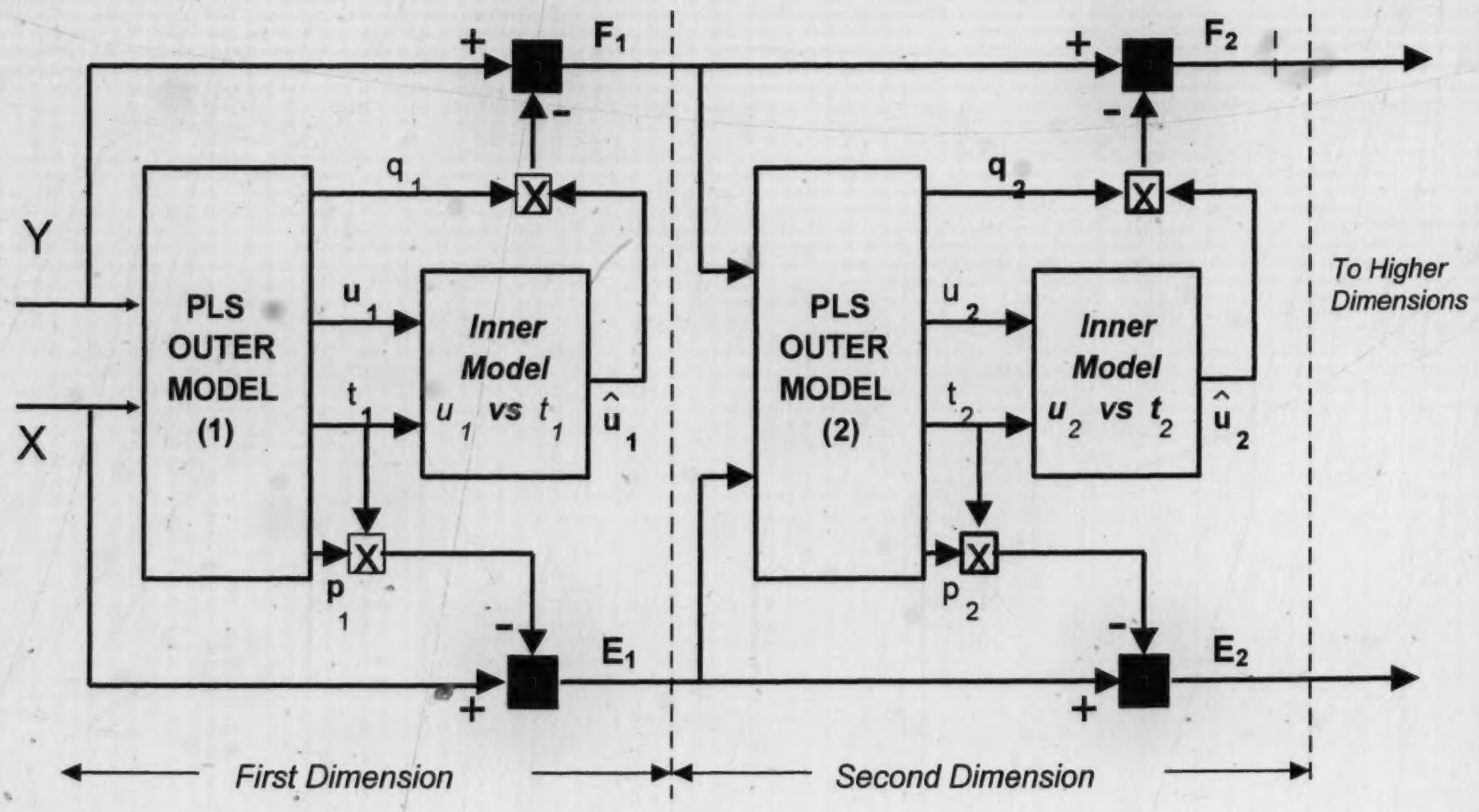


Figure 3-1: PLS Block Diagram

In order to get orthogonal X scores, slightly rotated X loading vectors are defined as

$$p_1 = X^T t_1 / \left(t_1^T t_1\right) \tag{3-5}$$

Y loadings q_1 are the same as Y-weights l_1. Geometrically, the loading vectors p_1 and q_1 are the direction cosines of the dominant directions within the data set and are obtained by maximizing the covariance between X and Y. The first set of scores vectors t_1 and u_1 are interpreted as the projection of the X and Y data respectively onto the loading vectors p_1 and q_1. The procedure of obtaining scores vectors t_1 and u_1 from X and Y is known as the PLS outer model and is depicted by the block "PLS OUTER MODEL (1)" in Figure 3-1.

Next, the matrices X and Y are indirectly related through their scores by an "Inner Model" which is just a univariate linear regression of t_1 on u_1

$$u_1 = t_1 b_1 + n_1; \quad \Rightarrow \hat{u}_1 = t_1 b_1 \tag{3-6}$$

The least square solution for b_1 is given by $u_1^T t_1 / \left(t_1^T t_1\right)$. The quantity $\hat{u}_1 q_1^T$ can be interpreted as the portion of the Y data that has been predicted by the first PLS dimension. In doing so, the $t_1 p_1^T$ portion of X data has been used up. Next, denoting $E_0 = X$ and $F_0 = Y$, the residuals are computed via a matrix deflation process (shown as dark squares in Figure 3-1):

$$E_1 = X - t_1 p_1^T = E_0 - t_1 p_1^T \tag{3-7}$$

$$F_1 = Y - \hat{u}_1 q_1^T = Y - b_1 t_1 q_1^T = F_0 - b_1 t_1 q_1^T \tag{3-8}$$

Successive PLS vectors (the scores and loading vectors and the inner relations) can be obtained for the deflated X and Y matrices. The procedure is continued (with the residuals computed at each stage) until all of the PLS dimensions (n) are extracted. The overall PLS algorithm is shown schematically in Figure 3-1.

In practice only a limited number of PLS dimensions (say r, usually r<<n) are required to give a reasonable fit between X and Y data. The number of PLS dimensions is determined based on the percentage variance explained, or by cross validation technique (explained in the PCA section). The directions considered irrelevant in the data sets,

which arise from noise and redundancies, are confined to the error matrices E_r and F_r. Thus the entire procedure can be seen as decomposing X and Y data into a sum of series of rank one matrices as follows:

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_r p_r^T + E_r = TP^T + E_r \tag{3-9}$$

$$Y = u_1 q_1^T + u_2 q_2^T + \cdots + u_r q_r^T + F_r = UQ^T + F_r \tag{3-10}$$

In order to obtain the PLS estimates of the parameter C of the linear model shown in Eq 3-1, we proceed from Eq 3-10 as $Y = UQ^T + F_r = TBQ^T + F_r$. Therefore, if r PLS dimensions are retained in the final model, prediction of Y is given by:

$$\begin{aligned}
\hat{Y} &= TBQ^T \\
&= [t_1 \mid t_2 \mid \cdots \mid t_r]BQ^T \\
&= [Xw_1 \mid E_1 w_2 \mid \cdots \mid E_{r-1} w_r]BQ^T \\
&= [Xw_1 \mid (X - t_1 p_1^T)w_2 \mid \cdots \mid E_{r-1} w_r]BQ^T \\
&= [Xw_1 \mid (X - Xw_1 p_1^T)w_2 \mid \cdots \mid E_{r-1} w_r]BQ^T \\
&= [Xw_1 \mid X(I_n - w_1 p_1^T)w_2 \mid \cdots \mid E_{r-1} w_r]BQ^T \\
&= X[w_1 \mid (I_n - w_1 p_1^T)w_2 \mid \cdots \mid E_{r-1} w_r]BQ^T \\
&= X[w_1^* \mid w_2^* \mid \cdots \mid w_r^*]BQ^T \\
&= XW^*BQ^T
\end{aligned}$$

Therefore, PLS estimate of C is given by

$$C_{PLS} = W^*BQ^T \tag{3-11}$$

Here $W^* \in \Re^{n \times r} = [w_1^* \mid w_2^* \mid \cdots \mid w_r^*]$ with $w_i^* = \prod_{h=1}^{i-1} (I_n - w_h p_h^T) w_i$, $B \in \Re^{r \times r}$ is a diagonal matrix containing b_i's as the diagonal elements, i.e., $B = diag[b_1 \quad b_2 \quad \cdots \quad b_r] \in \Re^{r \times r}$, and $Q \in \Re^{k \times r} = [q_1 \mid q_2 \mid \cdots \mid q_r]$ is the matrix of Y loadings.

The pioneering work in PLS was done by H. Wold in the late sixties (Wold, 1966). Since then, much literature has emerged to describe PLS theory and many algorithms have been proposed (Geladi and Kowalski, 1986; Hoskuldsson, 1988; Lindgren *et al.*, 1993; Dayal

and MacGregor, 1997). A tutorial description of PLS along with a simple example has been provided by Geladi and Kowalski (1986a, b). Manne (1987) and Hoskuldson (1988) provide an excellent analysis of the mathematical properties of the algorithm. The different PLS algorithms differ mainly in the iterative manner in which the eigenvectors are calculated. In this work, NIPALS (Non-Linear Iterative Partial Least Squares) algorithm (Geladi and Kowalski 1986a) was used to develop the basic PLS models. NIPALS is a robust procedure for solving eigenvalue eigenvector related problems where the eigenvectors are calculated in a partial fashion, one at a time, until all variance in the data structure is explained. For each new dimension, the information explained by the last component is subtracted from the data matrices X and Y to create residuals on which subsequent dimensions are calculated by the same procedure (Lindgren *et al.*, 1993). For practical application of the PLS algorithm, it may be necessary to scale the X and Y blocks suitably in view of the fact that the measurement units can be grossly different. Without proper scaling, the PLS latent variables may be significantly biased towards variables with larger magnitude. Scaling may be performed with some a priori knowledge, e.g. assigning larger weights to some key variables. Often all variables are auto scaled (mean centered and scaled to unit variance). The scaled X and Y blocks are then processed by the PLS algorithm. A summary of the NIPALS algorithm is given below.

PLS Algorithm:

1. Start with mean centered, and scaled X and Y; set $E_0 = X$; $F_0 = Y$
2. For each component: u_{start} = some Y_j
3. In the X block: $\left\{\begin{array}{l} w' = u'X / u'u \\ w'_{new} = w'_{old} / \|w'_{old}\| \\ t = Xw \end{array}\right\}$
4. In the Y block: $\left\{\begin{array}{l} q' = t'Y / t't \\ q'_{new} = q'_{old} / \|q'_{old}\| \\ u = Yq / q'q \end{array}\right\}$
5. Check for convergence
6. $p' = t'X / t't$

7. $p'_{new} = p'_{old} / \|p'_{old}\|$

8. $t_{new} = t_{old} \|p'_{old}\|$

9. $w'_{new} = w'_{old} \|p'_{old}\|$

10. Find regression coefficient b for inner relation: $b = u't / t't$

11. Calculate deflated X and Y: $\begin{Bmatrix} E_h = E_{h-1} - t_h p'_h; X = E_h \\ F_h = F_{h-1} - b_h t_h q'_h; Y = F_h \end{Bmatrix}$

12. Go to step 2 for the next component.

Once the PLS model is obtained using data collected from normal plant operations, it can be used for predicting the quality variables in an inferential framework. The PLS matrices (scores and loadings) can also be used for fault detection and isolation. It should be noted that the predictions provided by the PLS model are reliable as long as the plant-model mismatch is insignificant. In case of time-varying plants, it may be necessary to use recursive versions of the PLS algorithm.

3.3 Process Description

In this work, an industrial autoclave reactor at AT Plastics (Edmonton, Canada) was studied. The plant uses ICI high-pressure technology for polymerization of ethylene to LDPE and copolymerization of ethylene and vinyl acetate to EVA copolymers. The plant has been in operation for over five decades and is a major supplier of specialty polymer resins in North America. The data for this work was collected from the most recently built reactor unit (known as R-5, the fifth reactor) at the plant. The unit produces more than 35 grades of LDPE and EVA resins and is controlled using a TDC3000 Honeywell DCS. Figure 3-2 shows a simplified process flow diagram.

The four-zone reactor has a nominal capacity of 750L. Ethylene gas enters the reactor through four separate streams. Free radical initiators are injected continuously into each zone of the reactor to control the exothermic polymerization reaction at selected temperature levels. The reactor is equipped with an axial stirrer and thermocouples along the reactor length. The reactor operates at a pressure in the range of 1000 to 3000 atm and

achieves about 20% conversion in a single pass. After separation from the unreacted gases in the high pressure separator and extrusion hopper, molten polymer is fed into a single screw extruder which forces the polymer to an underwater pelletizing unit in which polymer is cut into solid pellets.

Ethylene gas separated from the high pressure separator is recycled through a series of coolers and cyclone separators to the suction of the secondary compressor. Cyclone separators are used to separate any low molecular weight polymer from the recycle gas. Any ethylene separated from the extrusion hopper is compressed in a booster compressor, and finally mixed with fresh ethylene at the suction of the primary compressor. Off-gas from the extrusion hopper is partially purged in order to control the accumulation of impurities in the system. The same unit also produces ethylene vinyl acetate copolymers (EVA) when vinyl acetate monomer is added to the system.

In addition to the major processing units shown in Figure 3-2, the polymer plant is also equipped with a train of heat exchangers. These are used primarily to cool various feed and product gas streams and can be broadly classified as follows:

- Inter-stage coolers: cools the feed gas as it passes from the 1st stage to the 2nd stage of the secondary compressor
- Feed gas cooler: cools the hot feed gas exiting from the secondary compressor before it enters the reactor
- Product cooler: cools the product stream from the reactor before it enters the high pressure separator
- Return gas cooler: cools the hot effluent gas from the high pressure separator before it enters the 1st stage of the secondary compressor

The recycle gas exiting from the high-pressure separator contains small amount of low molecular weight polymer. Upon cooling, these low melting point polymers are deposited as a solid scale on the internal surfaces of the tubes in the return gas coolers. The same phenomena take place in the suction and discharge piping of the secondary compressor,

the inter-coolers, feed gas coolers and product coolers. This results in increased pressure drop and a higher fouling factor, which affect conversion and plant throughput. In order to remove these scales from the tube surfaces, occasionally, the coolers are operated at temperatures significantly higher than the normal operating temperatures. At this condition the scale melts and is swept away with the feed or product gases, leaving the heat exchange surfaces clean. This procedure is known as cooler cooking and is performed as a routing maintenance operation in the plant during normal plant operation. Cooler cooking is also carried out during a number of grade transition operations. The product made during cooler cooking is off-grade material.

A trickle sampler is used to direct a small amount of polymer pellets to the quality control lab where quality parameters such as melt index, comonomer (vinyl acetate) content, film quality etc. are measured continuously. These measurements are used for taking instantaneous corrective actions required for maintaining grade specifications. Fundamental molecular properties such as MWD are measured in the product analysis lab for each batch of product. The plant uses an online rheometer (Gottfert ®) that measures MI continuously and the value is stored in a computer database called the historian. At steady state, this equipment provides an accurate measure of MI, but during grade transition, the estimates are often erroneous. For many grade changes, the equipment is taken offline to change the capillary die and no measurement is taken. In addition, due to an inappropriate calibration technique used to store data in the historian, all MI values are stored only as a two-digit number (see section 4.6 for detail). This made interpretation of the available MI data during large grade transitions extremely difficult. A manual analytical test is also performed in this lab at an irregular frequency to experimentally measure MI. These spot measurements are used to update the online measurements for any bias correction, and also during grade transition to make decisions about changing process conditions. Points worth mentioning about the analytical measurement include irregular sampling rate leading to many missing data and manual recording of the data on log sheet with reliable/unreliable time stamp. For these reasons, online MI values have been used in this work for building the softsensor after appropriate scaling.

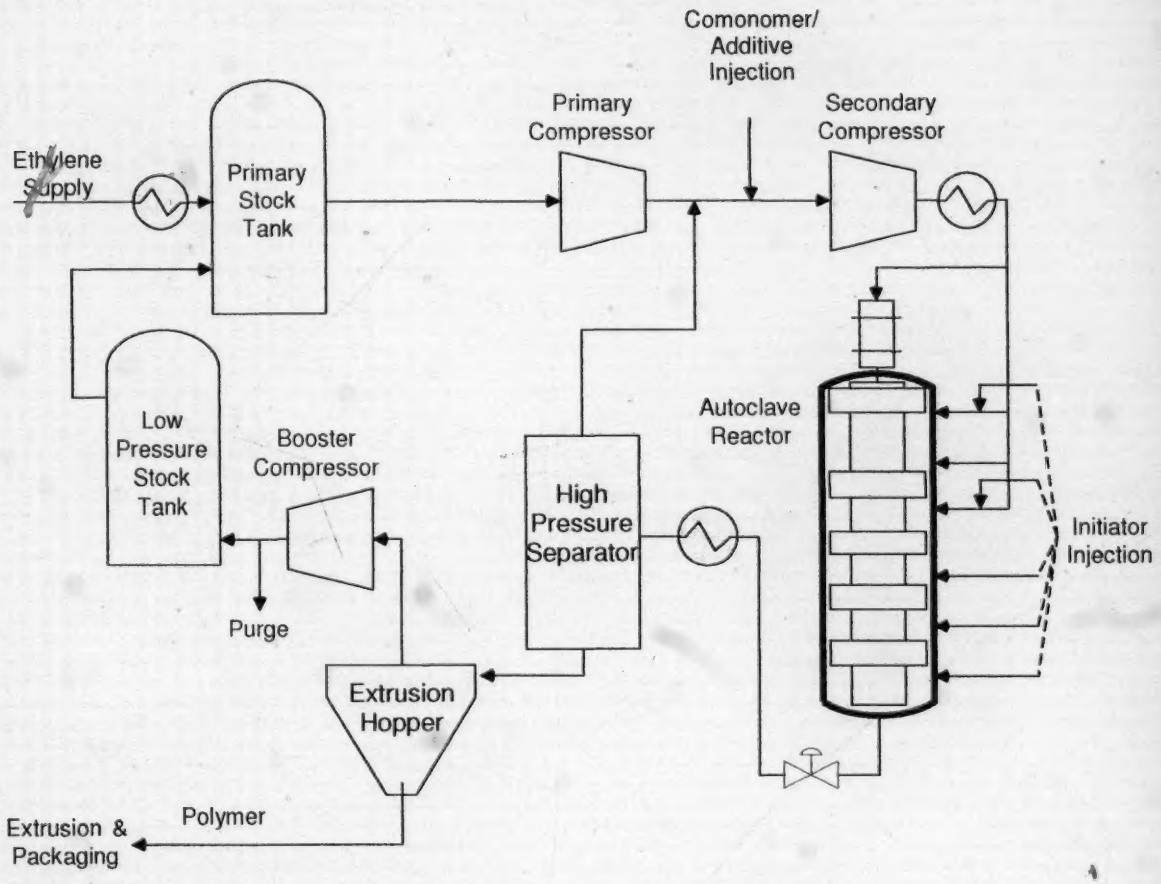


Figure 3-2: Polymerization process flow diagram

3.4 Grade Transition and Modeling strategy

To meet the current market demand for diverse products, while at the same time, trying to minimize the equipment cost, it is usual that most polymer plants will produce many grades of products by changing their reactor operating conditions. The common practice for product scheduling in most plants is as follows. Soon after the specification of the desired polymer is given by the customers, polymer quality that the individual processing unit produces is specified appropriately. Operating conditions are then determined for

each unit to meet the specified quality. Taking plant stability into account, a production schedule is determined so as to meet the demand while keeping the inventory cost low. Following the determined schedule, safe and quick changeover operations are performed by a sophisticated control system.

The complexity of operating procedures during grade change depends upon the differences in the product recipes. Product changeover usually consists of partially overlapping activities, i.e. phasing out the previous product, adapting and changing over to the new product, and developing analytical procedures to determine the new product quality. Generally, the grade transition is achieved by continuously changing the catalyst, modifiers and other operating conditions. In certain cases when the recipes are widely different, particularly with regard to the catalyst type, grade transition requires temporary reactor shutdown and then restarting. Process complexity is increased when another operation such as a cooler cook is done simultaneously with the grade change to minimize downgrade material. In the current project, the plant produces more than 26 grades of LDPE and EVA products with a melt index ranging from as low as 0.3 to as high as 2200. The vinyl acetate content in the EVA copolymer can vary from 16-40%. As a result of frequent grade transitions, the plant operation becomes increasingly complex. In summary, modeling of this process is a not a trivial task for the following critical factors (Piovoso and Pearson, 1992; Kumar, 2001)

- Multiple grades are manufactured by the same process unit based on current market demand
- Unscheduled changeovers or special operating procedures (such as cooler cooking) are used to revive operating efficiency
- Modifications are sometimes done to the standard operating procedure for a grade
- Effect of unplanned disturbances

In view of the above mentioned complexities, a decision had to be made about the type of the soft sensor model that would be suitable for such a diverse plant. The approach taken during this research project was to develop a number of composite models that would be

applicable over a range of melt index values. These models include both the steady state and grade transition regions. A number of single grade models were also developed. Only those grades which are manufactured frequently in the plant were selected for model building. It is worth mentioning that the appropriate length of the training data required to build any model depends on whether we intend to build a composite model or a model for single grade product. In both cases, plant production history was studied to select the duration and length of the data. The usual time length for producing any particular grade of polymer varies from 2~7 days at the plant. Several batches of data, each having length ranging from 2~7 days, was collected to build single grade models. For composite models, data over one month period were collected. In this work fast rate of data storage placed an upper limit on the length of the training data.

3.5 Data Acquisition and Preprocessing

Three basic steps in any data based modeling or monitoring scheme are data collection, data preprocessing and development of the model. Though it has received little attention in the literature, data preprocessing can be the key to the success or failure of the final application. Many publications on modeling and monitoring of industrial processes start with statements such as "Assume that the data is available in a matrix X". However, in most cases, data from an industrial process does not exist in a form that is suitable for use by softwares which are used to build monitoring and soft sensor applications. A large amount of data preprocessing is required in order to pre-treat the data and transform it in a usable form as is assumed in such statements. In reality, more than 80% of the time required to develop online monitoring and soft sensor applications is spent in data preprocessing and data quality analysis (Raghavan, 2004). The following sections summarize the procedure of data collection and the preprocessing steps followed in this work.

3.5.1 Data Collection

An OPC server (Object linking and embedding for Process Control) was used to access data from the Universal Control Network (UCN) of a Honeywell TDC3000 DCS. The OPC servers used Microsoft's OLE technology to communicate with clients and permits

standards for real-time data exchange between Windows based software applications and process control hardware. The data communication between the server and the DCS was via two RS-232 serial interface cards installed on one of the marshalling cabinets on the Universal control Network of the DCS. Offline process data for a total of 48 variables was queried from the server using Structured Query Language (SQL) program files. These included all major process variables from the reactor, and product quality variables such as Melt Index, Vinyl Acetate content (mass%), etc. Additional variables were collected from all other important processing units of the plant, which included variables from the compressors, product cooler, product separators, and the extruder sections. A list of these variables is given in Appendix-2. Data for all 48 variables were accessed at every second and saved in a database folder. As a standard practice in the process industry, special operations and events are logged by the plant personnel in a shift wise fashion (one shift is 12 hours). The same approach was found to be advantageous in data collection. Therefore, one data folder was created to store the data for one shift, which contains 48 separate files for each variable. However, the original process data was collected in a form, which was incompatible with the format required for any data processing software (i.e., MATLAB, Excel, or ProcessDoc). A sample representation of one of these data files as collected from the plant site is shown in Table 3-1. A MATLAB code was written to select only the numerical values from the 3rd column from each file and paste them into one large data matrix. Thus one data matrix was generated for each shift's operation. The columns of the matrix represented the variables and the rows represented the measurements at different sampling instants.

Table 3-1: Sample of the data as collected from the plant site

Tagname: TI51015.PV			
ip_trend_time		ip_trend_value	ip_trend_qstatus
01-MAR-02	06:00:00.2	192.160	Good
01-MAR-02	06:00:01.2	192.160	Good
01-MAR-02	06:00:02.3	192.160	Good
01-MAR-02	06:00:03.3	192.160	Good

3.5.2 Assessing Data Quality

Any data based modeling or monitoring task starts with the stored process data retrieved from the plant historian. Most industrial historians use some built-in data preprocessing functions for reasons such as minimizing data storage cost, improving the quality of the data from univariate display point of view etc. Some of these preprocessing steps introduce unwanted changes in the original data, which may alter the results obtained from multivariate analysis when we use the processed data. Moreover, the data may be corrupted with outliers and noise thus reducing the signal to noise ratio. Therefore, it is important to asses the quality of the data before performing tasks such as building static or dynamic inferential models or monitoring schemes using such data. Assessment of data quality is an important step which includes determination of the extent of data compression, estimating quantization effects, and outlier detection and smoothing. These steps are briefly described in the following sections.

Analysis of Data Compression: Process data stored in the historian can be a valuable source of information for activities such as development of inferential sensors, minimum variance control loop benchmarking, fault detection, data reconciliation etc. However, due to hardware and network bandwidth limitations it is difficult to store and retrieve large quantities of data. Therefore, in most plants, data compression is used to reduce the costs of storage of historical data and transmission of process data through a telecommunications link. A variety of data compression algorithms, many of them based on piecewise linear interpolation, are now being used by industrial historians. Some examples include box car algorithm, backward slope, swinging door algorithm etc. The swinging door algorithm (Inc, 2002) which is one of the most commonly used algorithms is illustrated in Figure 3-3. More detail on this method is available in Thronhill *et al.*, 2004.

Although data compression is useful from a data storage point of view, it has hidden costs if the data become unsuitable for their intended purposes. Thornhill *et al.* (2004) concluded that compression induces changes to many basic statistical properties of the data such as altering the mean and variances. It adversely affects the results obtained

from multivariate statistical analysis, process monitoring studies and inferences obtained from control loop performance assessment. A quantitative measure of compression present in any data is given by compression factor (CF) which is defined as follows:

$$CF = \frac{Number\ of\ samples\ in\ original\ signal}{Number\ of\ samples\ in\ compressed\ signal}$$

It has been demonstrated by Thornhill *et al.* (2004) that data with CF > 3 are not suitable for data driven analyses. Several techniques have been proposed by researchers to estimate the compression factor. In this work, the algorithm presented in Thornhill *et al.* (2004) was used. The results indicated very high compression in the data, with a minimum $CF_{min} = 6$.

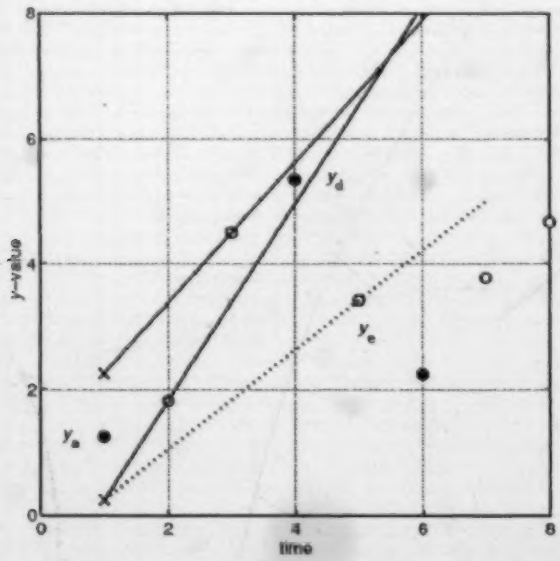


Figure 3-3: Illustration of swinging door compression. Black circles represent archived spot values. Values with open circles are not archived. At time step 5 (point y_e) the lower door (dotted line) opens up wider than parallel showing that a new trend started at y_d.

A thorough investigation of the historian's data storage procedure revealed that the historian only stores new data every 6 seconds and uses zero order hold for intermediate values. Process variables are sampled in the plant using a sampling time of either 1 second (for flow, pressure, and temperature) or 12 minutes (for composition measurements from GC). However, values are displayed on the DCS at every second using a zero order hold device. Due to technical limitation, the historian takes 6 seconds to collect all 48 tags. For example, in the first second the first 8 tags are accessed; in the next second, next 8 tags are accessed while the values for the first eight tags remain constant at their previous levels. In this way, in 6 seconds, all 48 tags are accessed and stored in the historian. In the seventh second, the historian goes back to measure the first set of 8 tags and the cycle repeats. Therefore, even for a variable that changes rapidly (such as flowrate), there will be many straight-line segments where at least six values will remain constant (see Figure 3-4), resulting in a minimum compression factor of six. For a slowly changing variable (composition, temperature), the length of the straight-line segments may be even larger. Any compression detection algorithm will falsely indicate the presence of data compression if this particular feature of data storage is ignored. Therefore, it is concluded that the data obtained was essentially compression free and can be used in any data based analysis. However, since the data is updated at every 6 seconds, this should be the minimum down sampling rate.

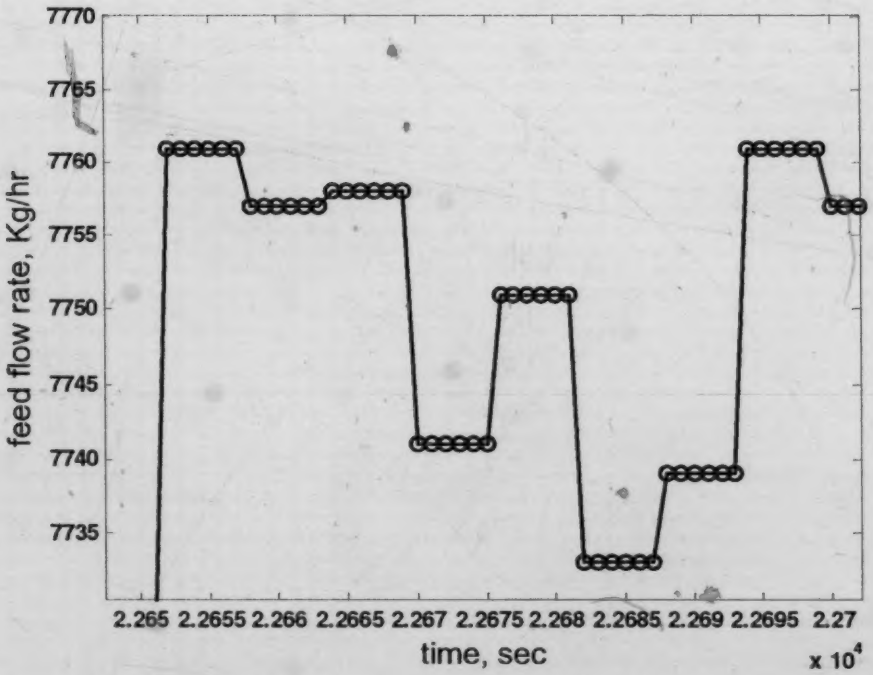


Figure 3-4: Plot of fresh feed flow rate shows how the data is stored in the historian

Data limits and quantization: During creation of a new tag in the historian, it is a common practice to define upper and lower limits on the values. If the physical variable exceeds these limits, the actual value is truncated to these pre-defined limits and is stored. In statistical literature this phenomena is defined as data censoring (Gupta, 1952). During data collection, care should be taken to note these limits for each variable of interest. These limits should be set to meaningful values such as the range of the measuring instrument.

Data quantization is defined as the minimum step size used for descretization in the data. Quantization results from two factors namely, the data storage allotment in the historian, i.e., the number of bytes allotted to each data point and the physical capabilities of the instrument. Data storage has become relatively inexpensive in recent days. Hence, data quantization due to storage limitation is generally not a big problem. However variables

might get quantized due to limitation of the measuring instrument. Figure 3-5 shows an example of highly quantized data originating from a gas chromatograph. In this case, quantization resulted from physical limitations of the instrument.

Before proceeding with further analysis, it is important to look for these effects in the data as it may affect the results of the analysis. It is safer to discard the tags which are subjected to high degree of quantization and censoring. If it is believed that the variable might have significant influence in the model building, then appropriate measures should be taken to compensate for these effects. For example, it might be necessary to use inferential strategies to predict the value of the sensor reading when it gets censored. The expectation maximization algorithm (Dempster *et al*, 1977) provides one such strategy. If one does find out that a particular variable behaves in a discrete fashion because of physical or storage limitation, it might be necessary to resort to discrete or hybrid regression techniques (Raghavan, 2004).

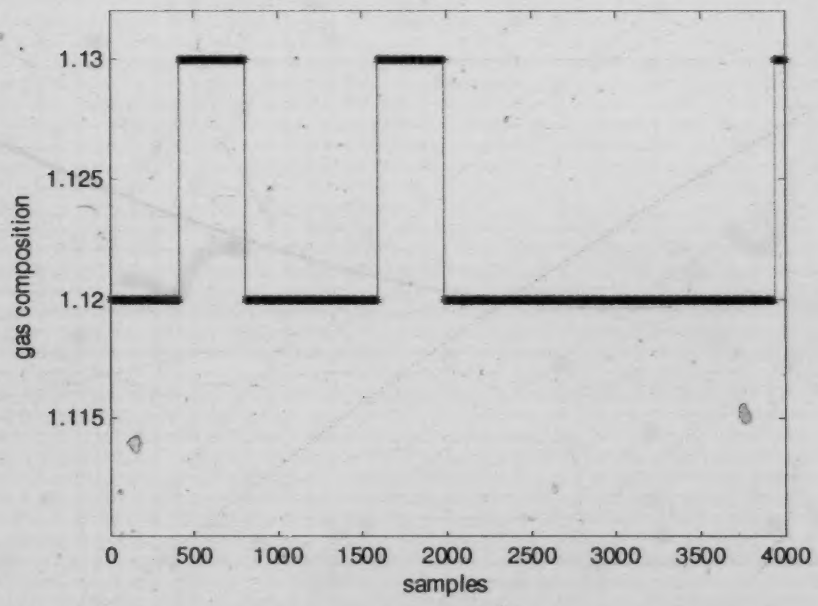


Figure 3-5: Example of data quantization

Outlier Detection and Smoothing: Process measurements are often corrupted with random noise which is typically concentrated in the high frequency regions of the signal. Low signal to noise ratio leads to reduced resolution and may cause false alarms during process monitoring. The same effect is observed if "bad data" or outliers are present in the measurements. Outliers are observations that do not follow the statistical distribution of the bulk of the data, and consequently may lead to erroneous results with respect to statistical analysis. A common method used to detect outliers can be represented as follows: a sample point x is regarded as an outlier if it satisfies $|x-t| > kv$, where t is a measure of data location (mean or expected value), v is a measure of the scatter (standard deviation), and k is a predefined value (k = 3, for 3-sigma method). Outliers may be generated by sensor fault, process fault, or human-related errors. Outliers can lead to model misspecification, biased parameter estimation, poor forecasts, and wrong analysis results and should be removed before any data based analysis. In this work, we used a revised Martin and Thompson (MT) filter-cleaner algorithm (Liu *et al.*, 2004) to detect and replace the outliers. The method includes an on-line outlier-resistant estimate of the process model and combines it with a modified Kalman filter to detect and "clean" outliers. This method was used for the following reasons:

- a priori knowledge of the process model is not required
- robust to autocorrelated and even nonstationary process data
- it tries to only clean (i.e., detects outliers and replaces them with expected values) outliers and preserves all other information in the data
- simple and easy to implement

The effect of random noise was reduced by filtering the data using an exponentially weighted moving average (EWMA) low pass filter that removes high frequency fluctuations. This filter can be represented by a 1st order transfer function as follows:

$$y(t) = x(t) - \tau_f \frac{dy(t)}{dt} \tag{3-12}$$

Here $x(t)$ is the raw measurement at time t, $y(t)$ is the filtered value at the same instant, and τ_f is filter time constant. Using backward difference to approximate the derivative in Eq 3-12, we write,

$$y_n = x_n - \tau_f \left(\frac{y_n - y_{n-1}}{\Delta t} \right) \tag{3-13}$$

After simplification, we get,

$$y_n = \alpha\, x_n + (1-\alpha) y_{n-1} \tag{3-14}$$

Here $\alpha = \left(1 + \frac{\tau_f}{\Delta t}\right)^{-1}$. The filter time constant τ_f is usually selected based on the frequency range of noise associated with the signal. It should be chosen such that it is much smaller than the dominant time constant of the process. In this application, the residence time in the reactor, which was close to 42 seconds for most grades of product, was chosen as the dominant time constant. Accordingly, a value of $\tau_f = 4$ was selected. Using this value of τ_f, Eq 3-12 was discretized and used as the model equation for the EWMA filter:

$$y_n = 0.85 x_n + 0.15 y_{n-1} \tag{3-15}$$

It should be noted that a higher value of τ_f gives more smoothing, which may be desirable if we want higher resolution. However, in this case we compromise on the prediction time of faults since a dynamic lag is introduced (Seborg and Mellichamp, 2004). Some fine tuning of the filter parameter was carried out based on the noise level of the data. Manual removal of some bad data and outliers was accomplished using the plant daily event log sheets where any special events, faults, etc. that occurred during the operation are recorded.

3.5.3 Sorting for In-line Pumps

Each zone of the reactor is supplied by its own catalyst injection pump and is backed by another back-up pump. Therefore, a total of eight catalyst pumps are used for this purpose, but only four of them are on line at any one point in time. Outputs from all eight pumps were collected in the raw data. Therefore, we needed to identify which pump was on line during each day's operation, and keep only the data corresponding to that pump.

3.5.4 Calculation of mass flow rate

It is expected that the total reactor feed should play an important role in affecting the product quality. However, this quantity was not measured at the plant. A first principles model to estimate the mass flow rate through the secondary compressor was provided by the plant personnel and was used successfully in previous work by Kumar *et al.* (2003). This model estimates the mass flow rate of the discharge gas from the secondary compressor using the compressor suction and the discharge conditions. The same model was used in this work to calculate the mass flow rate. This additional derived variable was appended to the last column of the original data matrix.

3.5.5 Time delay estimation

It is important in any data based modeling exercise to appropriately time shift the data according to the time delay between different processing units to maximize the correlation in the data matrix. Using mass flow rates, equipment dimensions, and piping dimensions, the time delay was estimated from the feed points to the MI measurements taken at the lab. The procedure for calculating time delay is summarized in Appendix-3. Some of the estimates are shown in Table 3-2.

Table 3-2: Estimated Transport Delay

Units	**Estimated delay (sec)**
Reactor	42
Reactor to product cooler	1
Product cooler	27
Product cooler to inlet of high pressure separator	1.2
High pressure separator	176
High Pressure separator to low pressure hopper	45
Extruder	262
Pelletizer to trickle sample collector	10
Lab Extruder	180

Finally, as mentioned earlier, for PLS modeling, it is essential to use some scaling method. In this application, the data was scaled by removing the mean and dividing by the standard deviation. Coefficients obtained from models based on both scaled and unscaled data indicated that, when the data was not scaled, variables with large numerical values heavily dominated the model. Variables with large numerical values in most cases also exhibit larger standard variations than variables with smaller numerical values of the same data set. Therefore, one should expect that PLS on unscaled variables is dominated by those variables.

3.6 Results from PLS Models

Because of the wide range of melt index covered by the various polymer grades, we initially selected two intervals, one from the high MI range and one from the low MI range, and built two composite models that would be valid for each interval. The final data matrices, after completion of all preprocessing steps, were used for model building. Variables that were collected from the fresh ethylene feed point up to the reactor were used as input to the PLS model. All setpoints and variables measured downstream from the reactor were discarded. The model structure is as follows:

$$\ln\{MI(t)\} = a_1 x_1(t - t_{d_1}) + a_2 x_2(t - t_{d_2}) .. a_n x_n(t - t_{d_n}) \qquad (3\text{-}16)$$

where x_1 to x_n are process variables, t_{d_i} is the time delay between sensor i and MI analyzer in the lab, and a_i are PLS regression coefficients. All models were written as MATLAB codes using MATLAB's PLS toolbox. It should be noted that results on model performances shown in this chapter and in all subsequent chapters in this thesis are based on validation data.

One composite model (high MI) was developed using data collected from four grades of polymer having melt index ranging from 6 to 110. Online steady state data was used for model building. Figure 3-6 shows the prediction from this model. The correlation coefficient between the actual and the predicted values was 0.99 indicating the goodness of fit. Another PLS model was developed by combining data for four polymer grades having low melt index (between 2 and 7). The model contains one output and 22 input

variables. Both steady state and transient data were used in model building and the data was appropriately time shifted using the estimated time delay. Figure 3-7 shows the prediction from this model. The correlation coefficient between the actual and the predicted values was 0.82. The spikes in the true MI values in Figure 3-7 represent large overshoot during grade transition. Also in some instants, product quality deviated significantly from target values due to some abrupt change in operating conditions. The model was able to capture some of these trends.

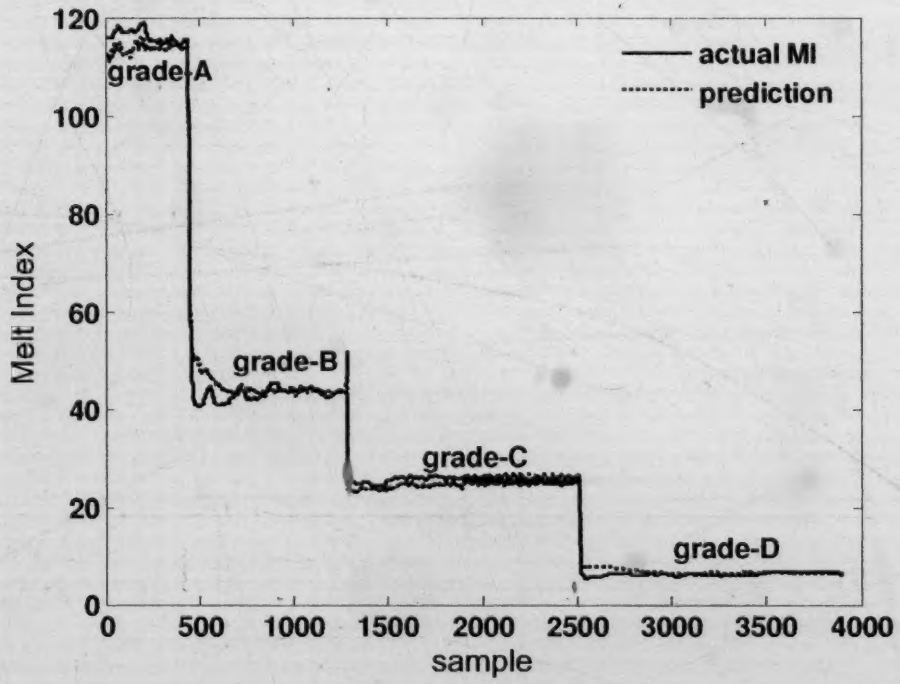


Figure 3-6: Prediction from high MI PLS model

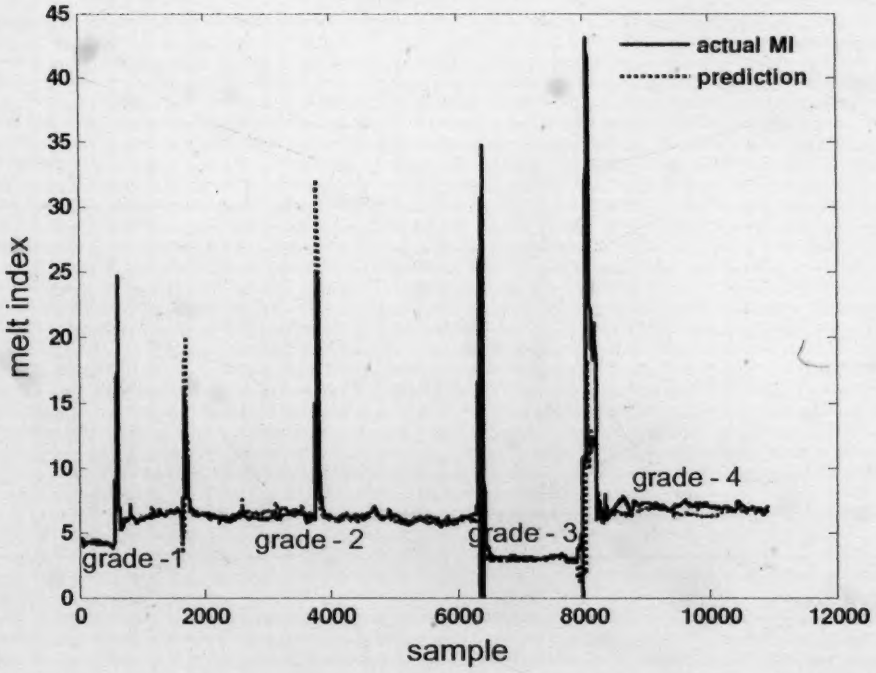


Figure 3-7: Prediction from low MI PLS model

It is important to note that although correlation coefficients between the actual and the predicted values are quite high over the entire range of MI, the predictive capability of these composite models are not as good for stationary data segments. For example, in Figure 3-7, if we focus on each grade separately, correlation coefficients are only 0.33, 0.88, 0.49, and 0.25 for grade 1, 2, 3, and 4 respectively. This implies that although the universal models capture large variation in MI between grades, small variation within a grade are not well captured. It should be noted that often, large overshoots in MI values occur during grade transitions. When steady state data are combined with transient data that contain large variability, prominent peaks during transition tend to mask smaller variation during steady state operation. This is the reason why the composite models show poor performance during production of the single grades. In order to overcome this deficiency, we built separate steady state models for each single grade. The universal

models would be used for grade transition only; and once steady state was reached, the appropriate single-grade model would be used for estimation of MI.

A single grade PLS model was built using the data for grade-D (see Figure 3-6). This EVA product has a target melt index of 6. Figure 3-8 shows the model prediction. 14 PLS dimensions captured 93.7% variance in X and explained 73% variance in Y. The correlation coefficient between the predicted and the true MI values for validation data set is 0.70 compared to 0.30 obtained from the composite model.

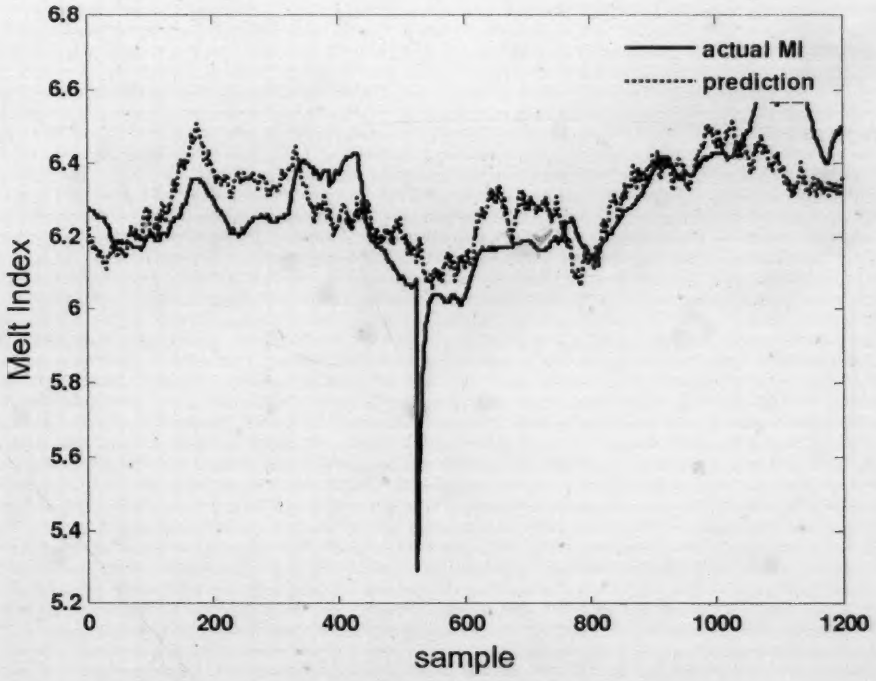


Figure 3-8: PLS model prediction for grade-D

Figure 3-9 shows model prediction for the PLS model developed using data from grade-C only (see Fig 3-6). The model captures the variation in melt index quite well, and results in a correlation coefficient of 0.7 when applied to validation data set.

Another single grade PLS model was developed for one specialized grade of EVA product with high MI (grade-E, not shown in Fig 3-6). The target melt index for the grade is 480. The plant operators sometimes find it difficult to maintain the MI within specification. We developed the softsensor to predict MI for this grade. Figure 3-10 shows the model prediction. The model with 25 PLS dimensions captured 99% variance in X and explained 94% variance in Y. It captures the timely variation in melt index quite well, and, when applied to validation data set, results in a correlation coefficient of 0.7, although there is a bias. It may be mentioned in this regard that significant variations in the process operating conditions were observed between different batches of data for this particular grade of product. This made it difficult to ensure that the training and the validation data sets had similar variability, which is required for good prediction ability of the PLS models. Moreover, since the target melt index is high, it increases the uncertainty associated with the measured melt index values. These factors resulted in large bias in the predicted melt index values.

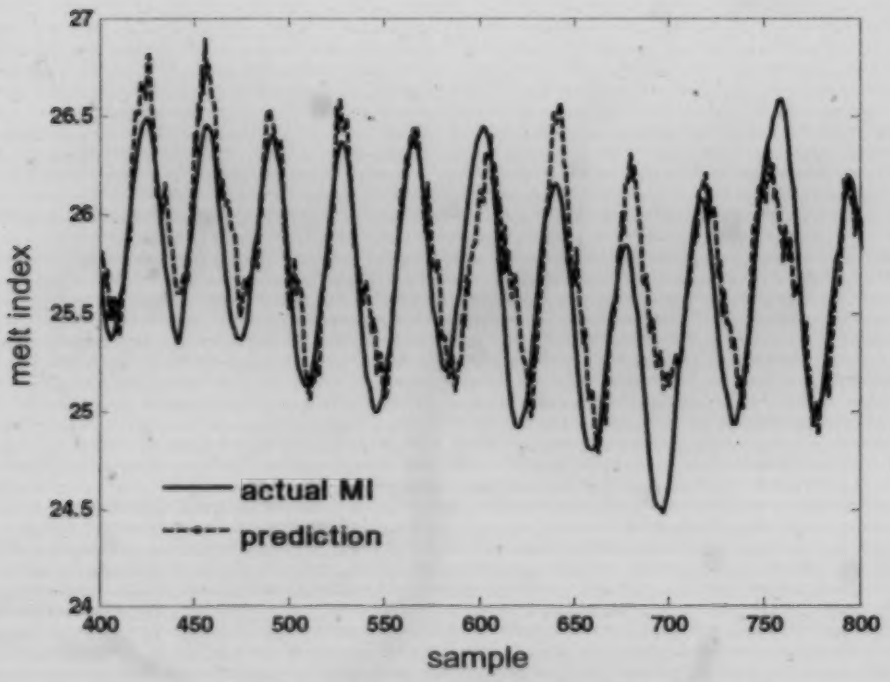


Figure 3-9: PLS model prediction for grade-C

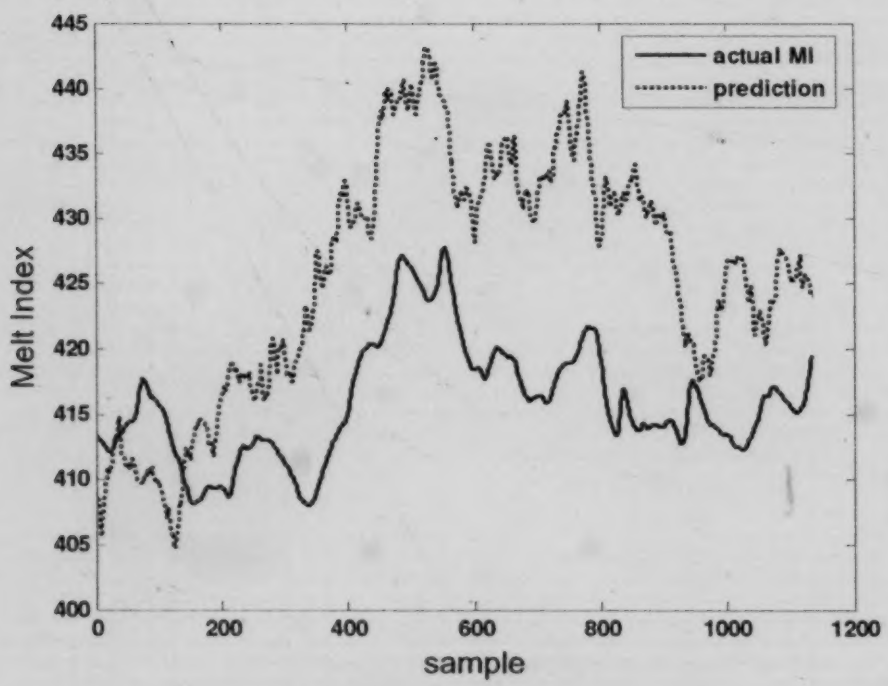


Figure 3-10: PLS model prediction for grade-E

3.7 Model Implementation: Need for Updating

Although the single grade PLS models showed acceptable performance for a number of grades of polymer, before implementing the models in real time, it was important to validate the models with data other than the training data set. A model built with a limited amount of data from one production campaign was used to predict properties during a different production campaign of the same grade of polymer. The model developed for Grade-D (in Fig 3-8) was applied to three data sets collected from three different production campaigns for the same grade. Note that the elapsed time between when the model was first built and tested and the new data sets are 3, 4, and 15 months respectively. Figure 3-11 shows the model prediction for these new data sets. It is clear from this figure that model performance deteriorates over time. It was observed that the variation in the data is significantly different at these different times. Examples of

changes in process variables with time are shown in Figure 3-12. For certain product recipes, the plant operators use the reactor pressure as a manipulated variable to control polymer properties. This can result in significant variation in pressure for different batches of data as is evident in Figure 3-12(a). To some extent, ethylene feed gas temperature depends on the ambient temperature and therefore shows seasonal variation as seen in Figure 3-12(b). Similar variations were noticed in a number of process variables for the three sets of data. Since PLS is a technique which is based on covariance among variables, it is important that the training data and the validation data have similar variability, i.e., the covariance structure of the data does not change over time. Therefore, it is important to update the model periodically to account for variation in the data (eg. ambient temperature variations at different times of the year may affect the variability in process conditions).

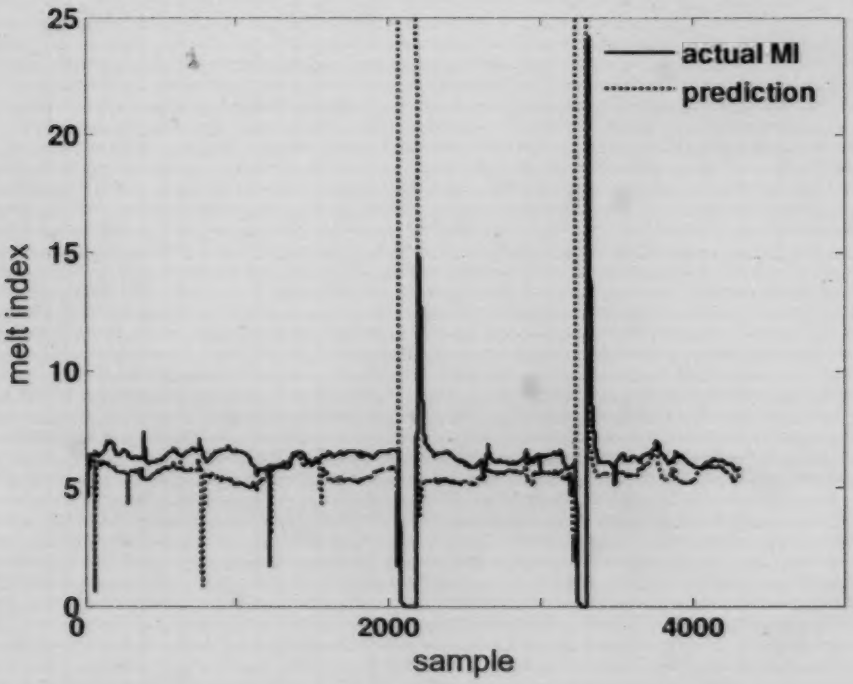


Figure 3-11(a): Model validation with three new data sets for Grade-D (after 3 months)

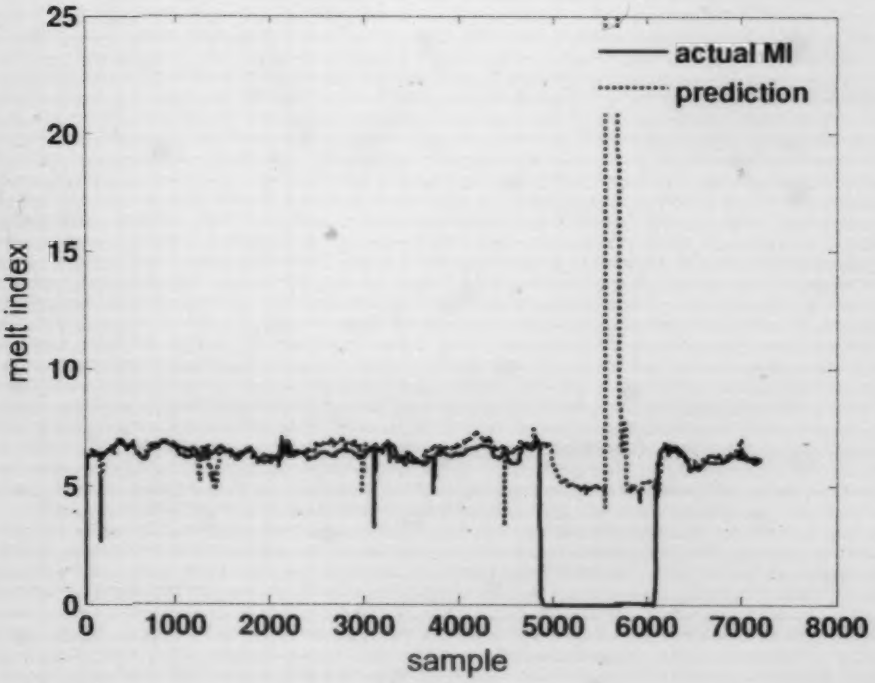


Figure 3-11(b): Model validation with three new data sets for Grade-D (after 4 months)

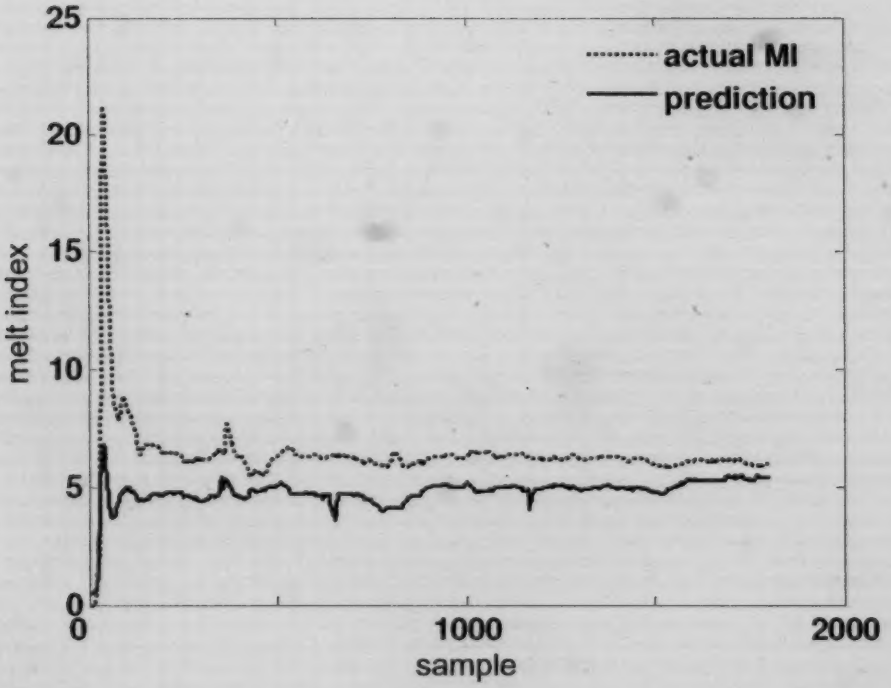


Figure 3-11(c): Model validation with three new data sets for Grade-D (after 15 months)

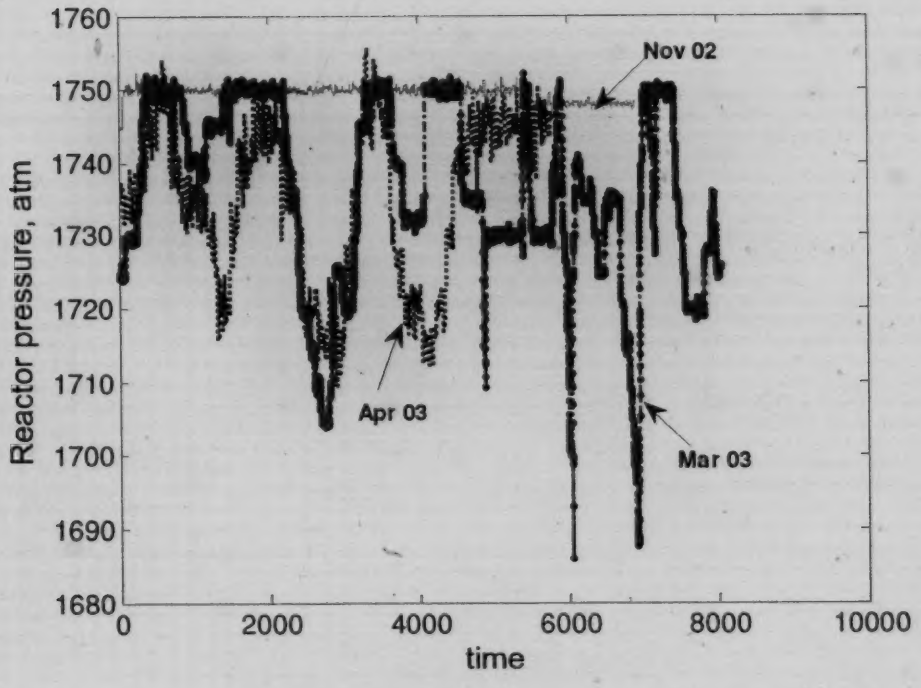


Figure 3-12(a): Variation in reactor pressure

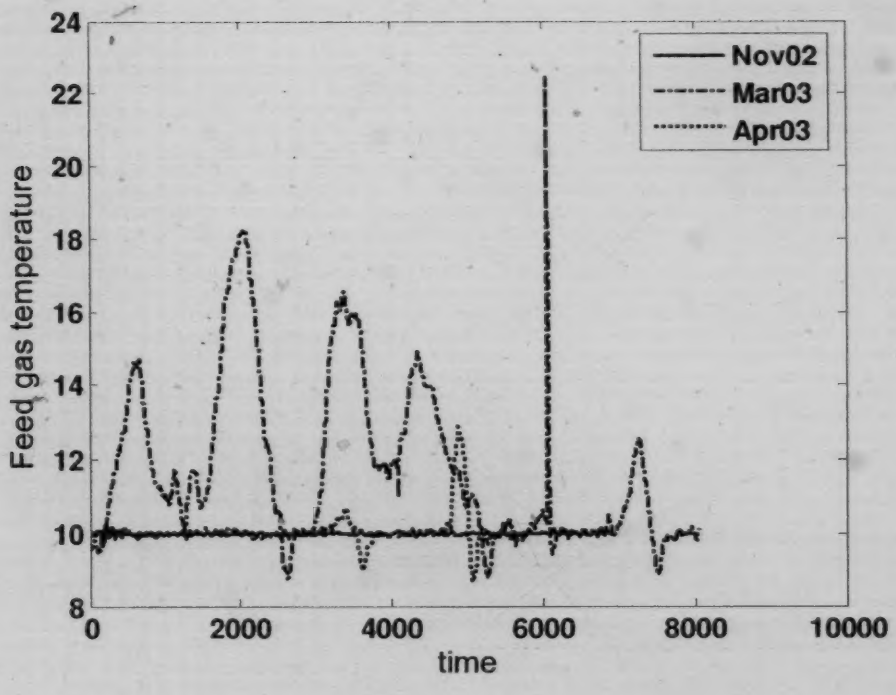


Figure 3-12(b): Variation in reactor feed gas temperature

Another point worth mentioning is that in all cases, validation data was scaled using mean and standard deviation of the training data. Therefore, the model that was applied for prediction was:

$$\left(\frac{Y-\bar{Y}_1}{S_Y}\right)=C_{PLS}\left(\frac{X-\bar{X}_1}{S_X}\right) \qquad (3\text{-}17)$$

Here X and Y are validation data set, $\bar{X}_1$, $\bar{Y}_1$ are the mean of the training data set, and S_Y, S_X are the standard deviation of the training data set. However, if the validation data were scaled using the mean of validation data and standard deviation of the training data, the following model would result:

$$\left(\frac{Y-\bar{Y}_2}{S_Y}\right)=C_{PLS}\left(\frac{X-\bar{X}_2}{S_X}\right) \tag{3-18}$$

Here $\bar{X}_2$ and $\bar{Y}_2$ are means of X and Y blocks of the validation data set. Using the mean of validation data in building the PLS model suggests that periodic updating of the mean value of process variables is required during implementing the models. This can be easily accomplished by using a bias updating scheme to correct for the changing mean value. An improved result was obtained with this revised scaling. Figure 3-13 shows a comparison of two scaling techniques for the first data set, which is taken after three months as showed in Figure 3-11(a).

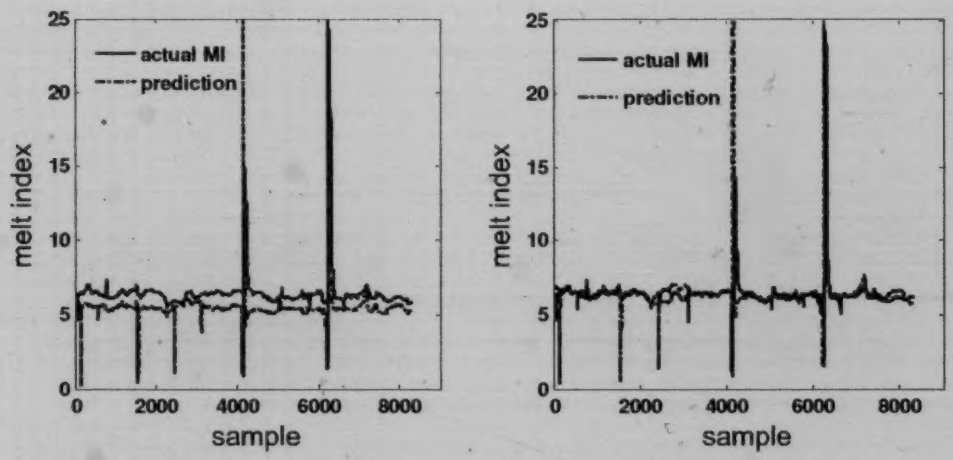


Figure 3-13: Model validation new data set for Grade-D: a) scaling factors taken from training data, b) scaling factor taken from validation data

It is evident that in order to successfully apply such models to a time-varying process, they must have some adaptive capability. Here we suggest a simple bias update scheme using spot values of melt index. Initially, the original PLS model is used for estimating MI. After a specified time interval, a bias term was calculated which is defined as the difference between the true MI value and the model prediction. This bias is added as a

correction term during the next specified time window. A feasible implementation strategy is as follows:

Let Y(t) represent the true MI value as measured using the rheometer. $\hat{Y}(t)$ represents the model prediction at time t. Therefore initially (at t = 0),

$$\hat{Y}(t) = X(t) \times b \tag{3-19}$$

At t = T,

$$bias(T) = Y(t) - X(t) \times b \tag{3-20}$$

For T < t < 2T,

$$\hat{Y}(t) = X(t) \times b + bias(T) \tag{3-21}$$

This procedure resulted in excellent prediction for all polymer grades. To illustrate an example, five new sets of data were collected for the production of Grade D. The elapsed time between when the model was first built and tested and the new data sets are 4, 5, 13, 14, and 15 months respectively. Figure 3-14 shows the prediction with and without the bias update.

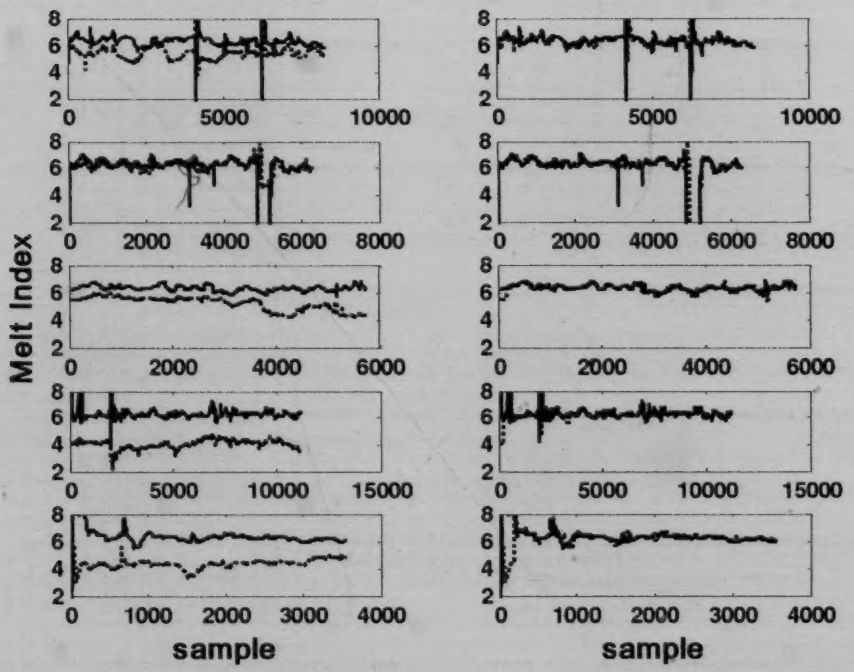


Figure 3-14: Model performance for Grade D after 4, 5, 13, 14, and 15 months (from top to bottom) without update (Left); with bias update at every hour (Right); (solid line: true MI, dashed line: model prediction)

Analytical samples of MI collected from the lab can also be used to update the bias correction on the output. However, irregular sampling in this case will result in less uniformity in the correction term. Nevertheless, the results obtained by updating the model using the lab samples are also promising as shown in Figure 3-15.

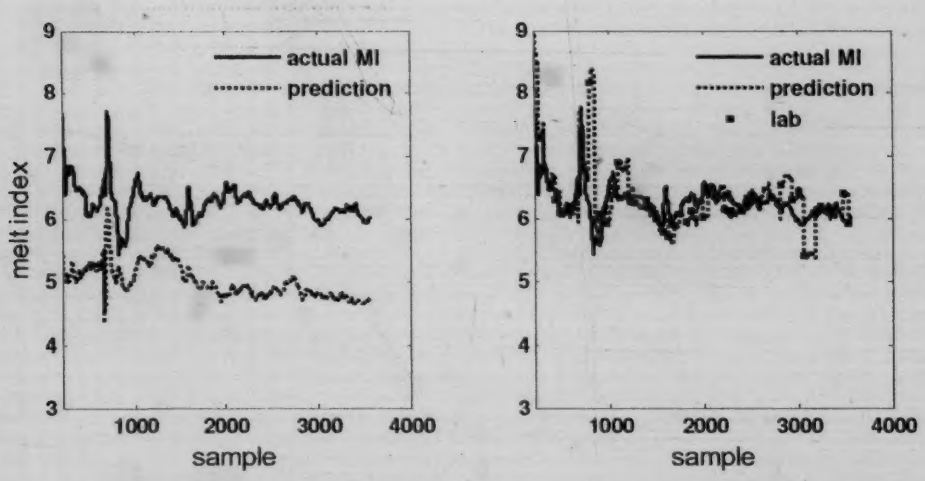


Figure 3-15: Prediction versus Lab MI samples: a) without bias update and b) with bias update

Models developed for two of the most frequent EVA grades were implemented at the plant in July 2006. The plant currently uses a Honeywell TDC 3000 DCS and an Aspen Tech's Infoplus21 historian to collect and store all the process data. A number of programs scripts were written using visual basics language to translate the MATLAB based model into an Aspen Tech Calculation script which is suitable for online use in the historian. Figure 3-16a and Figure 3-16b show the prediction from one of these models as compared to the MI values measured at the lab. The time difference between when the model was built and these e implementation results are 9 months (for Figure 3-16a), and 11 months (for Figure 3-16b). The model can quite successfully predict the actual MI using the bias update scheme. Even when the model shows minor deviation from the true value, the difference between the predicted and the actual MI are within acceptable range. The major advantage of using this soft sensor is that it predicts the MI values using the reactor conditions. The estimated delay between the reactor and the quality control lab is at least ten minutes. Therefore, current MI measurements at the plant (both online rheometer and analytical samples) are taken 10 minutes after the soft sensor predicts the value using data from the reactor. Therefore, the soft sensor will give the operator

reasonable time to take corrective action in case of a mismatch between the desired and the true MI values. This provides a powerful tool to perform product quality adjustments on-line just when there is an excursion in the reactor, resulting in better product for the customer, decreased material waste and reduced cost for the production plant.

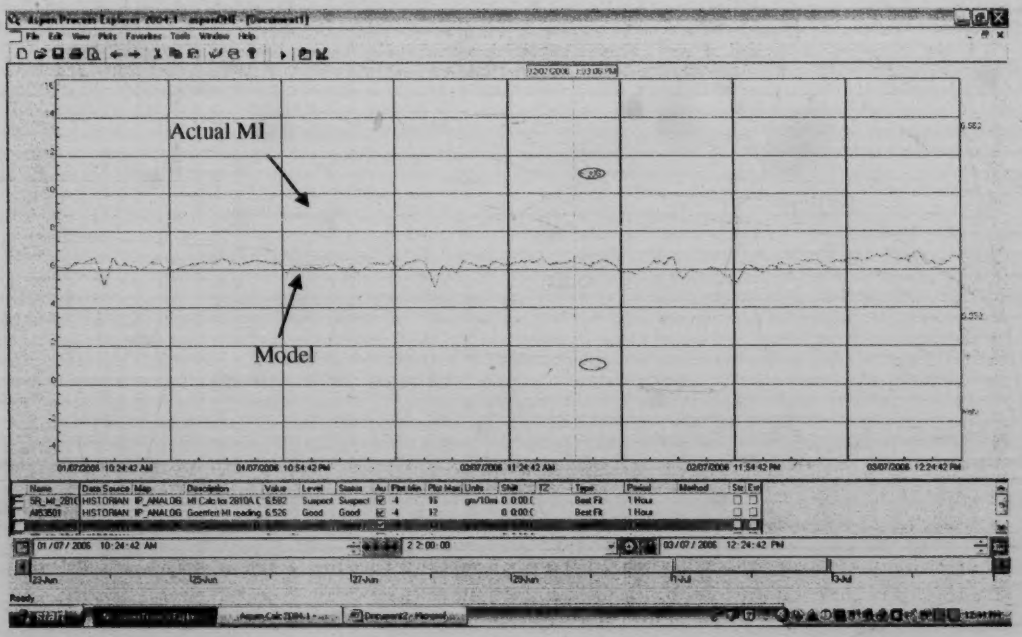


Figure 3-16a: Online implementation result for softsensor developed for grade D (July 2006)

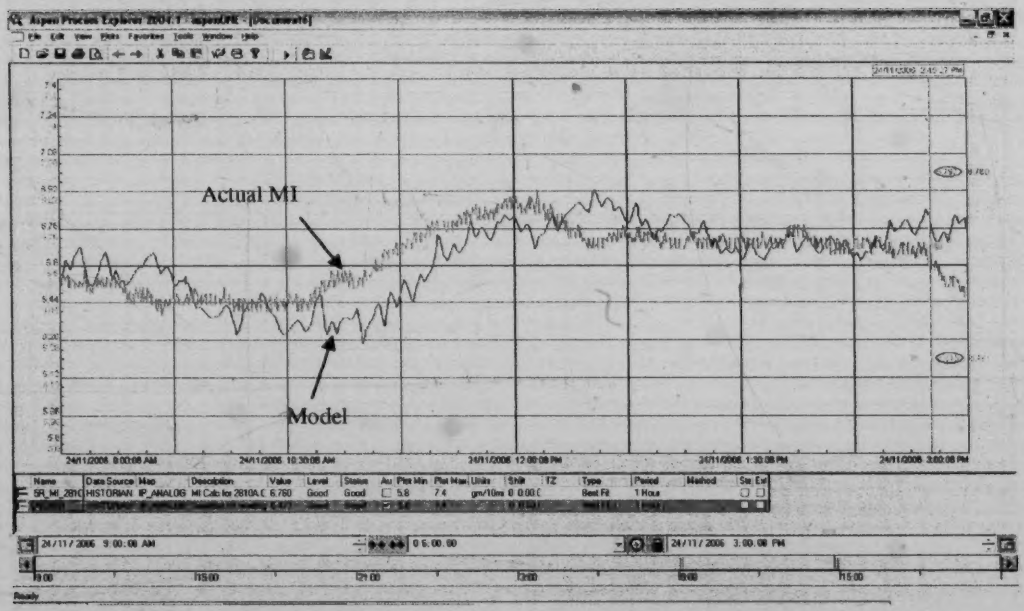


Figure 3-16b: Online implementation result for softsensor developed for grade D (Nov 2006)

3.8 Summary

In this chapter, we showed how PLS can be used to develop a soft-sensor to predict melt flow index using routinely measured process variables from an LDPE-EVA plant. Issues of data acquisition and preprocessing are discussed. The results indicated that melt index can be successfully predicted using this relatively simple statistical tool. It was observed that the model prediction becomes poorer with time as operators use new process conditions in the plant to produce the same resin with the same specification. Poor prediction results when the new conditions deviate significantly from the data with which the model was trained. We suggested a simple bias update scheme as a remedy to this problem. An alternative and more rigorous approach is to recursively update the model with new data, which will be described in the following chapter. Although we presented the method for modeling MI using PLS in this chapter, the method is general and can be applied to any other industrial data.

Chapter 4
Recursive PLS for Time Varying Processes

4.1 Introduction

Many industrial processes demonstrate time varying behavior due to reasons such as change of product grades, catalyst deactivation, equipment aging, sensor and process drifts, process changes incurred as a result of preventative maintenance and cleaning etc. (Gallagher *et al.*, 1997; Li *et al.*, 2000). A polymerization reactor is a typical example of a time variant system because of the multiple grades of polymer products obtained from the same reactor. During grade transition, reactor operating conditions are significantly changed over a course of time that can vary from two to six hours. Usually the polymer produced during this period is off specification and is sold at a lower cost in the market, reprocessed or discarded. In order to minimize the amount of off-specification product, it is important to accurately know the final quality of the polymer during this period so that operating conditions can be manipulated to move the process in the right direction. The main focus of this chapter is modeling of polymer property during grade transition.

One important property that is modeled in this work is melt index (MI). A data based PLS model for predicting properties, such as melt index, which is developed using steady state data collected from a single grade may be inadequate to explain the dynamic nature of the plant. A major limitation of such traditional steady-state PLS based modeling is that the model is time-invariant, while the real process is time-varying. The time-varying characteristics of industrial process data include: (i) changes in the mean; (ii) changes in the variance; and (iii) changes in the correlation structure among variables, including changes in the number of latent variables needed to model the system. When a time

invariant PLS model is used for prediction or monitoring of real processes, erroneous prediction and false alarms often result, which significantly compromise the reliability of the inferential or monitoring systems.

In order to accurately model a time varying process, adaptive modeling strategies are required. A number of recursive partial least squares (RPLS) modeling approaches have been published in the literature. The main idea of this adaptive modeling strategy is to initially build a static model using an initial set of data. Then, as the process changes, the initial model has poor capability in prediction or monitoring. Therefore, new information is "added" to the old model to build an updated model, which would then adequately represent the new operation conditions. This chapter presents a comparative review of all the existing RPLS techniques. A new RPLS method based on a novel interpretation of PLS using the so called controllability matrix is also presented.

The remaining part of this chapter is organized as follows: a brief summary of the recursive PLS theories is presented at the beginning; a detailed review of the RPLS algorithms follows; some examples of the dynamic nature of grade transition for a LDPE/EVA reactor is shown; and finally, application results from all RPLS algorithms are presented to model a EVA grade transition.

4.2 Recursive PLS Theories: an Overview

In many applications of data analysis, given a set of process variables, our objective is to relate a subset of those variables (known as response variables) to the rest of the process variables (known as explanatory variables). This is generally known as regression. Consider a process represented by the following regression model:

$$y_0(k) = C^T x_o(k) \tag{4-1}$$

Where $x_o(k) \in \Re^l$ and $y_o(k) \in \Re^m$ are the noise free explanatory variables and response variables respectively, at observation number k. $C \in \Re^{l \times m}$ represents the model parameter matrix. Assuming that a series of data is collected from the process described by Eqn (4-1) and the data are corrupted by noise, each measurement may be written as:

$$x(k) = x_o(k) + v(k) \tag{4-2}$$

$$y(k) = y_0(k) + w(k) \tag{4-3}$$

Where $v(k)$ and $w(k)$ are independent Gaussian distributed random noise vectors with zero mean and covariance matrices $S_v \in \Re^{l \times l}$ and $S_w \in \Re^{m \times m}$ respectively. Substituting Eqn (4-2) and (4-3) into Eqn (4-1) gives:

$$y(k) = C^T x(k) + e(k) \tag{4-4}$$

where $e(k) = w(k) - C^T v(k)$ accounts for the effects of measurement noise on the inputs and the outputs and follows a multivariate Gaussian distribution with zero mean and covariance $S_e \in \Re^{m \times m}$ given by $S_e = S_w + C^T S_v C$ (Johnson & Wichern, 1998). Using N observations k = 1,...,N, we can define an input $X \in \Re^{N \times l}$ and output $Y \in \Re^{N \times p}$ data matrix as follows

$$X = \begin{bmatrix} x^T(1) & x^T(2) & \cdots & x^T(N) \end{bmatrix}^T \tag{4-5}$$

$$Y = \begin{bmatrix} y^T(1) & y^T(2) & \cdots & y^T(N) \end{bmatrix}^T \tag{4-6}$$

$$E = \begin{bmatrix} e^T(1) & e^T(2) & \cdots & e^T(N) \end{bmatrix}^T \tag{4-7}$$

where the subscript 1 and N represent the time instants of the first and the last samples respectively. Accordingly, the relationship in Eqn (4-4) can be written as the following linear matrix equation:

$$Y = XC + E \tag{4-8}$$

A number of regression techniques such as ordinary least squares (OLS), principal component regression (PCR), partial least squares (PLS) etc. have been developed to get an estimate of C. Originally proposed by Wold (1966) for regression of linear time invariant (LTI) systems with correlated input variables, partial least squares regression has been widely applied in many fields for process modeling and monitoring. For example, PLS regression methods have been used in chemometrics (Lindberg *et al*, 1983; Wold *et al.*, 1984; Geladi and Kowalski, 1986; Fuller *et al.*, 1988; Martin and Naes 1989), in steady state process modeling (Piovoso and Owens, 1991), and in dynamic modeling (Ricker 1988; Wise and Ricker 1990; MacGregor *et al.*, 1991; Nomikos and

MacGregor 1995). PLS was also used in subspace (dynamic) system identification (Di Ruscio, 1997) to compute a basis for the observability matrix, which is the basis of most subspace identification algorithms.

In most of the reported applications, PLS has been used as a batch wise modeling approach. In other words, a batch of offline process data are collected, then the PLS regression is carried out on the whole batch of data. The model thus developed is then applied to predict future process outputs. While this approach is acceptable for static systems where the operating conditions do not change significantly, it will fail for dynamic or time varying processes. For such applications, it is therefore necessary to update the regression model using new data as they become available.

Recently a number of recursive algorithms for PLS have been published in the literature. These algorithms are developed as an extension of a basic PLS algorithm. As mentioned earlier, PLS was first developed as an iterative technique which involves calculation of eigenvalues and eigenvectors from the covariance matrix of the data. This is known as the NIPALS (Non linear iterative partial least squares) algorithm. In developing this theory, emphasis was given on numerical robustness and prediction capability of the model while less focus was given on computational efficiency. Helland *et al.* (1991) were the first to develop a recursive algorithm for PLS (RPLS), which was entirely based on the NIPALS algorithm. In their RPLS algorithm, an initial PLS model is built using a set of data in matrices X and Y. The model is updated at each new observation x_{new} and y_{new}. In the updated data matrix, which is used to build the updated model, X and Y are replaced by a compact representation using the X and Y scores calculated from the previous model. This keeps the size of the updated matrix constant. Qin (1993, 1998) later developed block-wise RPLS based on Helland's original RPLS algorithm and included a moving window and a forgetting factor adaptation scheme for online implementation.

With the advancement of computers and data storage facilities in the 1990's, PLS was being applied to large data matrices containing large number of samples of many

variables. Lindgren *et al.*, (1993) developed a faster PLS algorithm known as the kernel algorithm, which was particularly advantageous for handling large data sets. In the kernel algorithm, X-scores (w) are computed as the eigenvectors corresponding to the largest eigenvalue of the matrix X^TYY^TX using the power method or singular value decomposition (SVD). The remaining PLS vectors are then directly calculated from X^TX, X^TY and w. Subsequent PLS dimensions are obtained from the deflated covariance matrices X^TX and X^TY instead of deflated X and Y (as in the original NIPALS algorithm). Dayal (1996) and Dayal & MacGregor (1997a, 1997b) developed an exponentially weighted RPLS algorithm based on the kernel algorithm. Exponentially weighted RPLS method is based on the following covariance updating equations:

$$\left(X^TX\right)_t = \lambda_t\left(X^TX\right)_{t-1} + x_t^Tx_t \tag{4-9a}$$

$$\left(X^TY\right)_t = \lambda_t\left(X^TY\right)_{t-1} + x_t^Ty_t \tag{4-9b}$$

Here x_t and y_t are new observation at time t, $(X^TX)_t$ and $(X^TY)_t$ are updated covariance matrices at time t, and λ is a forgetting factor ($0 < \lambda \leq 1$). At each new sample, the covariance matrices are updated as in Eqn (4-9), and a new PLS model is developed from the updated matrices.

Recently, Di Ruscio (2000) has developed a novel PLS algorithm by using the Krylov controllability matrix (KCM) of process data. This novel algorithm completely avoids SVD and deflation operations on any matrix. In addition it has been shown that the KCM based PLS algorithm gives smaller prediction errors than the existing PLS algorithms for MIMO (multiple input multiple output) process. By extending Di Ruscio's work, this chapter proposes a recursive version of Di Ruscio's algorithm for MIMO time varying processes. Since neither SVD nor deflation is involved, the newly developed RPLS algorithm is mathematically feasible and computationally efficient.

In all the RPLS algorithms mentioned so far, the basic idea is to start with a process model built from a limited amount of historical data. As the process moves to a different operating region, new data are collected. Information content in the new data is added

with the old model and a model updating step is initiated. The updated model is then used to predict future output values. In the following sections, we present a detailed comparative review of all available RPLS algorithms.

4.3 RPLS based on NIPALS Algorithm

The common classical PLS algorithm is based on further development of the NIPALS (non-linear iterative partial least squares) method presented by Wold in 1966. NIPALS is a robust procedure for solving eigenvector-eigenvalue related problems where the eigenvectors (components, factors) are calculated in a partial fashion, one at a time until all variance in the data structure is explained. For each new dimension, the information explained by the last component is subtracted from the data matrices X and Y to create residuals, from which subsequent dimensions are calculated by the same procedure (Lindgren *et al.* 1993). Given input and output data matrices X and Y, we may construct a linear model of the form given by Eqn (4-8)

$$Y = XC + E \tag{4-8}$$

Here C and E are the coefficient and noise matrices respectively. If PLS is used to estimate C, X and Y are decomposed into bilinear terms:

$$X = t_1 p_1^T + E_1 \tag{4-10}$$

$$Y = u_1 q_1^T + F_1 \tag{4-11}$$

Where t_1 and u_1 are known as X scores and Y scores vectors. They are calculated as a linear combinations of X and Y such that $t_1 = Xw_1$, and $u_1 = Yj_1$. w_1 and j_1 are the corresponding weight vectors defining the linear combination. p_1 and q_1 are X and Y loading vectors. It has been proven that u, j, t, and w are eigenvectors corresponding to the largest eigenvalue of YY^TXX^T, Y^TXX^TY, XX^TYY^T, and X^TYY^TX respectively. In the NIPALS algorithm, estimation of these loading and scores vectors is presented as an eigenvalue-eigenvector problem. Eqn (4-10) and (4-11) represent the PLS outer model. The latent score vectors are then related by a linear inner model

$$u_1 = b_1 t_1 + r_1 \tag{4-12}$$

here b_1 is called the inner model coefficient which is determined from a univariate linear regression of u on t. After completing the calculation of the first set of vectors, which

represent the first PLS dimension, the second set of vectors are calculated by decomposing residuals E_1 and F_1. The procedure is repeated until all the factors are calculated. In the original algorithm, the weights vector w and the loading vector p are usually normalized to have unit length. Helland *et al.* (1991) and Qin (1998) have slightly modified the basic algorithm by normalizing t instead of w and p. The basic batch wise NIPALS algorithm (with this modification) is summarized in Table 4-1.

Table 4-1: Batch wise NIPALS algorithm for PLS

1. Scale X and Y to zero mean and unit variance
2. Initialize $E_0 =$ X, $F_0 =$ Y, and $h = 0$
3. Let $h = h + 1$, and $u_h =$ any column of F_{h-1}
4. Iterate until convergence:

$$w_h = \frac{E_{h-1}^T u_h}{u_h^T u_h}$$

$$t_h = \frac{E_{h-1} w_h}{\left\| E_{h-1} w_h \right\|}$$

$$q_h = \frac{F_{h-1}^T t_h}{\left\| F_{h-1}^T t_h \right\|}$$

$$u_h = F_{h-1} q_h$$

5. Calculate X loading $p_h = \frac{E_{h-1}^T t_h}{t_h^T t_h} = E_{h-1}^T t_h$

6. Find inner model $b_h = \frac{u_h^T t_h}{t_h^T t_h} = u_h^T t_h$

7. Calculate the residuals

$$E_h = E_{h-1} - t_h p_h^T$$

$$F_h = F_{h-1} - b_h t_h q_h^T$$

8. Return to step 3 until all factors are calculated.

Let us consider that the data pair X and Y has l inputs and p outputs with N samples. If the rank of X is r, then it can be proved (Qin, 1988) that

$$E_r = E_{r+1} = \cdots = E_l = 0 \tag{4-13}$$

This result indicates that the maximum number of factors that is retained in a PLS model does not exceed r. It was also shown by Qin (1998) that the output residual F_i is orthogonal to the scores of the previous factors t_h, i.e. $t_h^T F_i = 0$, for $i \geq h$. A recursive version of PLS which is based on the algorithm shown in Table 4-1 is developed by utilizing these two specific properties of the residual matrices E and F.

Using an initial set of data X and Y, one can derive a PLS model using the algorithm given in Table 4-1, where number of PLS dimensions calculated in the model is equal to the rank of X. The model is represented as follows:

$$\{X, Y\} \xrightarrow{PLS} \{T, W, P, B, Q\} \tag{4-14}$$

Where

$$T = [t_1, t_1, \cdots, t_r]$$
$$W = [w_1 w_2, , \cdots, w_r]$$
$$P = [p_1, p_1, \cdots, p_r]$$
$$B = diag\{b_1, b_1, \cdots, b_r\}$$
$$Q = [q_1, q_1, \cdots, q_r]$$

Therefore, X and Y can be written as

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_r p_r^T + E_r = TP^T + E_r = TP^T \tag{4-15}$$

$$\begin{aligned} Y &= u_1 q_1^T + u_2 q_2^T + \cdots + u_r q_r^T + F_r \\ &= t_1 b_1 q_1^T + t_2 b_2 q_2^T + \cdots + t_r b_r q_r^T + F_r = TBQ^T + F_r \end{aligned} \tag{4-16}$$

By minimizing the squared residuals $\|Y - XC\|^2$, we have $X^T XC = X^T Y$. From this, we can derive an expression for the PLS regression coefficient

$$C_{PLS} = \left(X^T X\right)^+ X^T Y \tag{4-17}$$

Where $(.)^+$ denotes the generalized inverse defined by the PLS algorithm. An explicit expression of the PLS regression coefficient is given by

$$C_{PLS} = W^* B Q^T \tag{4-18}$$

where $W^* = \left[w_1^*, w_2^*, \cdots, w_r^*\right]$, and $w_i^* = \prod_{h=1}^{i-1}\left(I_l - w_h p_h^T\right) w_i$. When a new data pair $\{X_1, Y_1\}$ is available, we are interested in updating the PLS model using the augmented data matrices $X_{new} = \begin{bmatrix} X \\ X_1 \end{bmatrix}$ and $Y_{new} = \begin{bmatrix} Y \\ Y_1 \end{bmatrix}$. The resulting PLS model is

$$C_{new}^{PLS} = \left(\begin{bmatrix} X \\ X_1 \end{bmatrix}^T \begin{bmatrix} X \\ X_1 \end{bmatrix}\right)^+ \left(\begin{bmatrix} X \\ X_1 \end{bmatrix}^T \begin{bmatrix} Y \\ Y_1 \end{bmatrix}\right) \tag{4-19}$$

Since the columns of T are orthonormal, the following relation can be derived using the results in Eqn (4-15) and (4-16):

$$X^T X = PT^T TP^T = PP^T \tag{4-20}$$

$$X^T Y = PT^T\left(TBQ^T + F_r\right) = PT^T TBQ^T + PT^T F_r = PBQ^T \tag{4-21}$$

Therefore the expression for the PLS coefficient becomes,

$$C_{new}^{PLS} = \left(\begin{bmatrix} P^T \\ X_1 \end{bmatrix}^T \begin{bmatrix} P^T \\ X_1 \end{bmatrix}\right)^+ \left(\begin{bmatrix} P^T \\ X_1 \end{bmatrix}^T \begin{bmatrix} BQ^T \\ Y_1 \end{bmatrix}\right) \tag{4-22}$$

By comparing Eqn (4-19) and (4-22) the following theorem can be derived:

Theorem 4-1: Given a PLS model $\{X,Y\} \xrightarrow{PLS} \{T,W,P,B,Q\}$, and a new data block $\{X_1, Y_1\}$, performing PLS regression on data pair $\begin{bmatrix} P^T \\ X_1 \end{bmatrix}, \begin{bmatrix} BQ^T \\ Y_1 \end{bmatrix}$ results in the same regression model as performing PLS regression on the data pair $\begin{bmatrix} X \\ X_1 \end{bmatrix}, \begin{bmatrix} Y \\ Y_1 \end{bmatrix}$.

Proof of theorem 4-1 can be found in Qin (1998). This theory is the basis for the NIPALS based recursive PLS method. Instead of using the old and the new data to update the model, the RPLS algorithm updates the model using the old model and the new data. A block wise RPLS algorithm with a moving window approach is summarized in Table 4-2.

Table 4-2: NIPALS based block wise RPLS algorithm

1. Formulate the data matrices {X, Y}. Scale the data to zero mean and unit variance or as otherwise specified with a set of weights.
2. Derive a PLS model using algorithm given in Table 4-1: $\{X,Y\} \xrightarrow{PLS} \{T,W,P,B,Q\}$. Carry out the algorithm until $\|E_r\| \le \varepsilon$ (ε is the error tolerance).
3. Select window size N, collect next N samples. During the collection, continue to use the initial model for prediction. The data collected during N samples are X_1 and Y_1.
4. Scale X_1 and Y_1 in the same way as in step 1.
5. Formulate updated data matrices $X = \begin{bmatrix} P^T \\ X_1 \end{bmatrix}, Y = \begin{bmatrix} BQ^T \\ Y_1 \end{bmatrix}$
6. Build PLS model using updated X and Y. Use this new model for prediction in the next window of samples.

4.4 RPLS based on Kernel Algorithm

The kernel algorithm for PLS was developed due to a growing need for a faster PLS algorithm to deal with large data matrices. The algorithm, which is a collective result of the research work by Joreskog & Wold (1982), Hoskuldsson (1988), and Lindgren *et al.* (1993) is now known as a faster alternative to the NIPALS algorithm. Applying traditional PLS to large data matrices often becomes demanding in terms of both memory requirement and computational time. The Kernel algorithm presents an alternative method to calculate all the crucial parameters in a PLS model using a small kernel (this will be defined later) and covariance matrices.

One common objective of using PLS is usually to understand the influence of X on Y and to develop a model for predictive purposes. The final result from the traditional PLS algorithm (such as NIPALS) is a set of weights (w, j) and loadings (p, q) which may be converted into a set of regression coefficients for predictive purposes. An expression for the PLS regression coefficient is given by:

$$C = W(P^T W)^{-1} Q^T \tag{4-23}$$

It is evident from Eq. 4-23 that the regression coefficient can be estimated using only the PLS weights and loading matrices, and that the scores are not needed for its estimation. Therefore it is theoretically possible to estimate PLS regression coefficients without calculating any scores. Since the size of the loading and the weight matrices are considerably smaller than the scores, a faster PLS algorithm will result by avoiding calculation of any scores. This is the basis for developing the kernel algorithm for PLS. This approach is advantageous for dealing with matrices with large number of observations as compared to variables.

It has been mentioned earlier that the vectors w, q, t, and u can be determined as eigenvectors corresponding to the largest eigenvalues of the following matrices:

$$w\alpha_1 = \left(X^T Y Y^T X\right) w \tag{4-24}$$

$$q\alpha_2 = \left(Y^T X X^T Y\right) q \tag{4-25}$$

$$t\alpha_3 = \left(XX^T YY^T\right)t \tag{4-26}$$

$$u\alpha_4 = \left(YY^T XX^T\right)u \tag{4-27}$$

The square matrices (YY^TXX^T, Y^TXX^TY, XX^TYY^T, and X^TYY^TX) shown in Eqn (4-24) to (4-27) are known as kernel matrices. Note that the kernel matrix X^TYY^TX in Eqn (4-24) has the smallest size$(l \times l)$. Therefore eigenvector-eigenvalue decomposition of this matrix using the power method or using singular value decomposition (SVD) will be the least computationally intensive. Once the weight vectors w are determined, the loadings required in estimating the PLS coefficient can be calculated through some simple manipulation of the variance (X^TX) and the covariance (X^TY) matrices. In estimating the q and p matrices using the original NIPALS algorithm (refer to PLS algorithm in section 3-2), the following relationships are used:

$$\text{X score: } t = Xw \tag{4-28}$$

$$\text{X loading: } p = \frac{t^T X}{t^T t} \tag{4-29}$$

$$\text{Y loading: } q = \frac{t^T Y}{t^T t} \tag{4-30}$$

Substituting Eqn (4-28) in (4-29) and (4-30) results in the following relationships:

$$p^T = \frac{w^T\left(X^T X\right)}{w^T\left(X^T X\right)w} \tag{4-31}$$

$$q^T = \frac{w^T\left(X^T Y\right)}{w^T\left(X^T X\right)w} \tag{4-32}$$

Therefore the first PLS set of factors (w_1, p_1, q_1) are calculated from Eqn (4-24), (4-31), and (4-32) using only X^TYY^TX, and X^TX, and X^TY. For estimating the next PLS dimension, these matrices must be updated by substituting the deflated E_1 matrix in X. Updating Y is not necessary for computing a PLS regression solution as was shown by Hoskuldsson (1988). Deflation of X is accomplished by:

$$E_1 = X - t_1 p_1^T = X - X w_1 p_1^T$$

$$\therefore\ E_1 = \left(I - w_1 p_1^T\right) X \tag{4-33}$$

Eqn (4-33) shows that a $I \times I$ square matrix $\left(I - w_1 p_1^T\right)$ updates X to E_1. Using the rules of matrix multiplication, it is possible to update $X^T Y Y^T X$, $X^T X$, and $X^T Y$ for the next dimension by direct multiplication. The general equation for updating becomes:

$$\left(X^T Y Y^T X\right)_{a+1} = \left(I - w_a p_a^T\right)^T \left(X^T Y Y^T X\right)_a \left(I - w_a p_a^T\right) \tag{4-34}$$

$$\left(X^T X\right)_{a+1} = \left(I - w_a p_a^T\right)^T \left(X^T X\right)_a \left(I - w_a p_a^T\right) \tag{4-35}$$

$$\left(X^T Y\right)_{a+1} = \left(I - w_a p_a^T\right)^T \left(X^T Y\right)_a \tag{4-36}$$

This is the end of the algorithm for the first dimension. The updated matrices are used for estimating the next dimension. For each new dimension, the weight and the loading vectors w_a, q_a, and p_a are stored in separate matrices W, Q, and P. Finally these matrices are used to estimate the PLS regression coefficient according to Eqn (4-23). The complete kernel algorithm is given in Table 4-3.

Table 4-3: Kernel Algorithm for PLS

1. Start with mean centered and scaled data X, and Y, and set a = 1
2. Compute the kernel and covariance matrices

$$\left(X^T X\right)_a = X^T \times X$$
$$\left(X^T Y\right)_a = X^T \times Y$$
$$\left(X^T Y Y^T X\right)_a = \left(X^T Y\right)_a \times \left(X^T Y\right)_a^T$$

3. Determine w_a as the eigenvector corresponding to the largest eigenvalue of $(X^TYY^TX)_a$ using power method
4. Calculate p_a, and q_a using Eqn (4-31) and (4-32)
5. Update the kernel and covariance matrices using Eq (4-34), (4-35), and (4-36)
6. Store w_a, q_a, and p_a in W, Q, and P matrices and set $a = a + 1$
7. Go back to step 3 for the next dimension

The important feature of this algorithm is that the original data matrices X and Y are only used once to formulate the kernel and the covariance matrices. These later matrices are of much smaller size as compared to X and Y. Furthermore, deflation of X and Y is replaced by updating X^TYY^TX, X^TX, and X^TY, which is accomplished by direct matrix multiplications. In the calculation of the regression coefficients, the scores vectors are not required and are not calculated. Lindgren *et al.* (1993) showed in a simulated application that these features made the kernel algorithm much faster than the NIPALS algorithm.

In later work, De Jong and Ter Braak (1994), and Dayal and MacGregor (1997) proposed some modifications to the original kernel algorithm. The original algorithm took advantage of the fact that deflation of Y is optional; hence the kernel theory was based on deflation of only the X matrix. Later, it was shown in their work that either X or Y needs to be deflated in the kernel algorithm such that:

$$\left(X^T Y\right)_{a+1} = X_a^T Y_{a+1} = X_{a+1}^T Y_a \qquad (4\text{-}37)$$

Based on this result the original kernel algorithm was modified slightly. In the modified algorithm, only X^TY was deflated instead of deflating both X^TX and X^TY. Using the

definitions of the loadings given in Eqn (4-31) and (4-32), and through some simple algebraic manipulations, the deflation of X^TY can be simplified to:

$$(X^TY)_{a+1} = (X^TY)_a - p_a q_a^T (t_a^T t_a) \tag{4-38}$$

By comparing Eqn (4-38) with Eqn (4-36), it can be seen that the computational effort in the deflation step in the original algorithm is reduced by avoiding the multiplication of X^TX and X^TY with $(I - w_a p_a^T)$. Note that the kernel matrix X^TYY^TX and covariance matrix X^TX are formulated only once and are not deflated anywhere in the procedure. Moreover, updating X^TY using Eqn (4-38) involves calculation of the X scores t. In order to accommodate the undeflated X^TX and X^TYY^TX, and to facilitate the computation of t from the original X data, some further modifications were proposed. We can define t according to the original NIPALS theory as:

$$\begin{aligned} t_1 &= X_1 w_1 = X w_1 \\ t_2 &= X_2 w_2 = X(I - w_1 p_1^T) w_2 \\ &\vdots \\ t_A &= X(I - w_1 p_1^T)(I - w_2 p_2^T) \cdots (I - w_{A-1} p_{A-1}^T) w_A \end{aligned} \tag{4-39}$$

Further we can define

$$\begin{aligned} r_1 &= w_1 \\ r_2 &= (I - w_1 p_1^T) w_2 \\ &\vdots \\ r_A &= (I - w_1 p_1^T)(I - w_2 p_2^T) \cdots (I - w_{A-1} p_{A-1}^T) w_A \end{aligned} \tag{4-40}$$

Note that the columns of the R matrix can be sequentially computed from the following recursive relationships:

$$\begin{aligned} r_1 &= w_1 \\ r_i &= w_i - p_1^T w_i r_1 - p_2^T w_i r_2 - \cdots - p_{i-1}^T w_i r_{i-1} \end{aligned} \tag{4-41}$$

This results in further improvement in computational efficiency of the modified algorithm as compared to computing r using Eqn (4-40). Now the score vectors T can be directly computed from the original X as follows:

$$T = XR \tag{4-42}$$

$$[t_1 \quad t_2 \quad \cdots \quad t_A] = X[r_1 \quad r_2 \quad \cdots \quad r_A] \tag{4-43}$$

Finally, if X is not being deflated, then Eqn (4-31) and (4-32) in step 4 of the kernel algorithm (Table 4-3) need to be modified so that the loading vectors p_a and q_a can be computed using the non-deflated or original X-matrix. The complete algorithm is summarized in Table 4-4. The modified kernel algorithm was shown to be much faster than the original algorithm.

Based on the modified kernel algorithm Dayal and Macgregor (1997a, 1997b) proposed an exponentially weighted recursive version of PLS theory. The algorithm starts with a block of process data X and Y, from which a PLS model is calculated using the kernel algorithm. Once a pair of new observations x and y become available, the covariance matrices are updated according to the following relationships:

$$\begin{aligned} \left(X^T X\right)_t &= \lambda_t \left(X^T X\right)_{t-1} + x_t^T x_t \\ \left(X^T Y\right)_t &= \lambda_t \left(X^T Y\right)_{t-1} + x_t^T y_t \end{aligned} \tag{4-9}$$

Here x_t and y_t represent the new observation at time t, $(X^TX)_t$ and $(X^TY)_t$ are updated covariance matrices at time t, and λ is a forgetting factor ($0 < \lambda \leq 1$). The magnitude of λ determines how fast the old data are discounted by the new model. At each new sample, the covariance matrices are updated as in Eqn (4-9), and a new PLS model is developed from the updated matrices. The steps involved in implementation of a sample wise RPLS algorithm are shown in Table 4-5.

Table 4-4: Modified Kernel algorithm for PLS

1. Start with mean centered and scaled data X, and Y, and set a = 1

2. Compute the kernel and covariance matrices

$$\left(X^T X\right)_a = X^T \times X$$
$$\left(X^T Y\right)_a = X^T \times Y$$
$$\left(X^T Y Y^T X\right)_a = \left(X^T Y\right)_a \times \left(X^T Y\right)_a^T$$

3. Compute w_a as the eigenvector corresponding to the largest eigenvalue of $\left(X^T Y Y^T X\right)_a$

4. Compute r_a as follows:

$$r_1 = w_1$$
$$r_a = w_a - p_1^T w_a r_1 - p_2^T w_a r_2 - \cdots - p_{a-1}^T w_a r_{a-1} \quad for\ a > 1$$

5. Compute p_a and q_a:

$$p_a^T = \frac{r_a^T \left(X^T X\right)}{r_a^T \left(X^T X\right) r_a}$$

$$q_a^T = \frac{r_a^T \left(X^T Y\right)}{r_a^T \left(X^T X\right) r_a}$$

6. Deflate $\left(X^T Y\right)_{a+1} = \left(X^T Y\right)_a - p_a q_a^T \left(t_a^T t_a\right)$

7. Store
$$W = \left[w_1 \quad w_2 \quad \cdots \quad w_a\right]$$
$$P = \left[p_1 \quad p_2 \quad \cdots \quad p_a\right]$$
$$Q = \left[q_1 \quad q_2 \quad \cdots \quad q_a\right]$$
$$R = \left[r_1 \quad r_2 \quad \cdots \quad r_a\right]$$

8. Set $a = a + 1$ and go to step 2 for next dimension

9. After computing all latent factors, regression coefficient $C_{PLS} = W\left(P^T W\right)^{-1} Q^T = RQ^T$

Table 4-5: Sample wise RPLS based on modified kernel algorithm

1. Collect mean centered and scaled data block X and Y
2. Compute C_{PLS} using the modified kernel algorithm
3. At each new observation x_t and y_t, update the covariance matrices $(X^T X)_t$ and $(X^T Y)_t$ as shown in Eqn (4-9)
4. Let $X^T X = (X^T X)_t$ and $X^T Y = (X^T Y)_t$
5. Compute a new set of C_{PLS} using updated covariance matrices

As a final comment on the modified kernel algorithm, we see that although the deflation of X is not explicit in the theory, it is implicit in the calculation of the score vectors. Instead of using the original X in calculating higher scores ($t_2, t_3, \ldots, t_A$), residuals in X space are used. These residuals are calculated from a set of vectors ($r_2, r_3, \ldots, r_A$). Comparing Eqn (4-40) with Eqn (4-33) we may conclude that computation of r is equivalent to deflation of X.

4.5 KCM based PLS Theory and RPLS Extensions

Since the introduction of PLS in the early sixties, PLS has been presented in the literature as an iterative algorithm, i.e. partial or piece-wise linear regression. The PLS solution is usually presented in terms of the score vectors, loading vectors, weighting vectors, and various iterative orthogonalization (deflation) processes, in addition to the solution for the matrix of regression coefficients. In a recent study, Di Ruscio (2000) has interpreted the PLS algorithm from a very different perspective. It was shown that the basic PLS algorithm is non-iterative and can be computed as the optimal solution to a prediction error minimization problem. There exists a very simple and non-iterative algorithm for computing the PLS solution in terms of some weighting vectors only.

In this section, we will present a brief description of the PLS theory based on Krylov controllability matrix. We start with a brief review of some basic definitions from linear

algebra which are essential in the development of this theory. Later we present a recursive PLS algorithm based on the controllability based PLS theory.

4.5.1 Eigenvalue, Eigenvector & Cayley Hamilton Theorem

A real or complex number λ is called an eigenvalue of the $n \times n$ real matrix A if there exists a nonzero vector x such that

$$Ax = \lambda x \tag{4-44}$$

Any nonzero vector x satisfying Eqn (4-44) is called a (right) eigenvector of A associated with eigenvalue λ. In order to find the eigenvalues of A, we write,

$$Ax = \lambda x = \lambda I x \tag{4-45}$$

$$(A - \lambda I)x = 0 \tag{4-46}$$

Here I is a $n \times n$ identity matrix. Eqn (4-46) is a homogenous equation. If the matrix $(A - \lambda I)$ is nonsingular, then the only solution of Eqn (4-46) ia x = 0. Therefore in order to obtain a nonzero solution for x, $(A - \lambda I)$ must be singular or have a determinant of zero. We define the determinant of $(A - \lambda I)$ as

$$\Delta(\lambda) = \det(A - \lambda I) \tag{4-47}$$

Eqn (4-47) is a monic polynomial of degree n with real coefficients and is called the characteristic polynomial of A. Eigenvalues of A are obtained as the solution of the polynomial equation as follows:

$$\Delta(\lambda) = \det(A - \lambda I) = 0 \tag{4-48}$$

By expanding Eqn (4-47), we obtain

$$\Delta(\lambda) = \det(A - \lambda I) = \lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_{n-1}\lambda + \alpha_n \tag{4-49}$$

The Cayley Hamilton theorem states that a matrix satisfies its own characteristic polynomial. Therefore, according to this theory,

$$\Delta(A) = A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1}A + \alpha_n I = 0 \tag{4-50}$$

4.5.2 Krylov/Controllability Matrix

In linear algebra, the Krylov matrix K(x, A, j) is defined as

$$K(x, A, j) = \begin{bmatrix} x & Ax & A^2x & \cdots & A^{j-1}x \end{bmatrix} \tag{4-51}$$

In the control literature, the Krylov matrix is known as the controllability matrix. Let us consider the n-dimensional p-input state equation for a linear time invariant system:

$$\dot{x} = Ax + Bu \tag{4-52}$$

Where A and B are $n \times n$ and $n \times p$ real constant matrices. For this system, the $n \times np$ matrix defined in Eqn (4-53) is known as the controllability matrix.

$$C = [B \quad AB \quad A^2B \quad \ldots \quad A^{n-1}B] \tag{4-53}$$

The system defined by (A,B) is said to be controllable if C is full row rank (i.e., rank(C) = n), hence the name controllability matrix.

Di Ruscio (2000) pointed out that there exists a relationship between the PLS weighting matrix w and a so called Krylov matrix. It is known that the problem of computing many orthogonal decompositions have an equivalent problem of computing subspaces for a Krylov matrix. Correspondence with Krylov matrices and orthogonal decompositions are pointed out in Golub and Van Loan (1986). Krylov subspaces and PLS are also discussed in Helland (1988).

4.5.3 Krylov controllability matrix based PLS algorithm

Consider an input data matrix $X \in \Re^{N \times l}$ and output data matrix $Y \in \Re^{N \times p}$ consisting of N samples of l inputs and p outputs. Let us define the variance and the covariance matrices as follows:

$$\Sigma_X = \frac{X^T X}{N}, \; \Sigma_{XY} = \frac{X^T Y}{N} \tag{4-54}$$

Using the definitions of the Krylov matrix given in (Eqn 4-51), we can formulate a Krylov controllability matrix (KCM) for the matrix pair $\{\Sigma_X, \Sigma_{XY}\}$ as follows:

$$K_l = \begin{bmatrix} (\Sigma_X)^{l-1}\Sigma_{XY} & (\Sigma_X)^{l-2}\Sigma_{XY} & \cdots & (\Sigma_X)\Sigma_{XY} & \Sigma_{XY} \end{bmatrix} \in \Re^{l \times lp} \tag{4-55}$$

For the system represented by Eqn (4-8), if Σ_X has a rank l, i.e Σ_X is full column rank, the least square solution of C is given by,

$$C_{LS} = (\Sigma_X)^{-1}\Sigma_{XY} \tag{4-56}$$

Since Σ_X is a $l \times l$ real matrix, its characteristic polynomial is given by

$$\lambda^l + \alpha_1\lambda^{l-1} + \cdots + \alpha_{l-1}\lambda + \alpha_l = 0 \tag{4-57}$$

Where $\lambda > 0$ is an eigenvalue of Σ_X, and α_i's are coefficients of the characteristic polynomial. Consequently it follows from the Cayley Hamilton theorem that

$$\Sigma_X{}^l + \alpha_1\Sigma_X{}^{l-1} + \cdots + \alpha_{l-1}\Sigma_X + \alpha_l I = 0$$

$$\text{or, } (\Sigma_X)^l + \sum_{j=1}^{l}\alpha_j(\Sigma_X)^{l-j} = 0 \tag{4-58}$$

In Eqn. (4-58), $\alpha_j \neq 0$, j = 1,..., l is the jth coefficient of the polynomial. $(\Sigma_X)^0 = I_l$, where I_l denotes an $l \times l$ identity matrix. It turns out from rearrangement of Eqn. (4-58) that

$$I = -\frac{1}{\alpha_l}\left[\Sigma_X{}^l + \alpha_1\Sigma_X{}^{l-1} + \cdots + \alpha_{l-1}\Sigma_X\right]$$

$$\text{or, } I_l = -\frac{1}{\alpha_l}\left[(\Sigma_X)^l + \sum_{j=1}^{l-1}\alpha_j(\Sigma_X)^{l-j}\right]$$

$$= \sum_{i=1}^{l}\alpha_l^*(i)(\Sigma_X)^{l+1-i} \tag{4-59}$$

Where $\alpha_l^*(i)$ is the ith element of $\alpha_l^* = -\frac{1}{\alpha_l}\begin{bmatrix}1 & \alpha_1 & \cdots & \alpha_{l-2} & \alpha_{l-1}\end{bmatrix}^T$ for $i \in [1, l]$. Hence multiplying Eqn (4-59) by $(\Sigma_X)^{-1}$ leads to

$$(\Sigma_X)^{-1} = \sum_{i=1}^{l}\alpha_l^*(i)(\Sigma_X)^{l-i} \tag{4-60}$$

Eventually substituting Eqn. (4-60) in Eqn. (4-56) results in

$$C_{OLS} = (\Sigma_X)^{-1}(\Sigma_{XY}) = \sum_{i=1}^{l}\alpha_l^*(i)(\Sigma_X)^{l-i}(\Sigma_{XY})$$

$$\therefore C_{OLS} = K_l \begin{bmatrix} \alpha_l^*(1) I_p \\ \vdots \\ \alpha_l^*(l) I_p \end{bmatrix} = K_l \left(\alpha_l^* \otimes I_p \right) \qquad (4\text{-}61)$$

Where $K_l = \in \Re^{l \times lp}$ is the Krylov controllability matrix of pair $\{\Sigma_X, \Sigma_{XY}\}$ as defined in Eqn (4-55), and $\otimes$ is the Kronecker tensor product. Therefore the ordinary least square solution C_{OLS} can be expressed in terms of the controllability matrix of the pair $\{\Sigma_X, \Sigma_{XY}\}$ and the coefficients of the characteristic polynomial. This solution is based on the assumption that Σ_X is non-singular.

If Σ_X is singular, its inverse no longer exists. For this case, a PLS solution can not be written in terms of K_l as in this case this matrix will be rank deficient. However as proved by Di Ruscio (2000), one can still use a truncated Cayley Hamilton series to obtain an optimal approximation of $(\Sigma_X)^{-1}$. Let us denote the rank of Σ_X as R(Σ_X). We can choose an integer l_0 such that $1 \le l_0 \le R(\Sigma_X) < l$. Now define a reduced controllability matrix K_{l_0} for the matrix pair $\{\Sigma_X, \Sigma_{XY}\}$ as follows:

$$K_{l_0} = \left[(\Sigma_X)^{l_0-1} \Sigma_{XY} \quad (\Sigma_X)^{l_0-2} \Sigma_{XY} \quad \cdots \quad (\Sigma_X) \Sigma_{XY} \quad \Sigma_{XY} \right] \in \Re^{l \times l_0 p} \qquad (4\text{-}62)$$

Now, using l_0, we can write a truncated Cayley Hamilton series similar to Eqn (4-58).

$$\Sigma_X{}^{l_0} + \alpha_1 \Sigma_X{}^{l_0-1} + \cdots + \alpha_{l_0-1} \Sigma_X + \alpha_{l_0} I = 0$$

$$\text{or, } (\Sigma_X)^{l_0} + \sum_{j=1}^{l_0} \alpha_j (\Sigma_X)^{l_0-j} = 0 \qquad (4\text{-}63)$$

Rearranging Eqn (4-63) as was done in deriving Eqn (4-60), an estimate for $(X^TX)^{-1}$ can be obtained from $\sum_{i=1}^{l_0} \alpha_{l_0}^*(i) (\Sigma_X)^{l_0-i}$. Here $\alpha_{l_0}^*(i)$ is the ith element of vector $\alpha_{l_0}^* \in \Re^{l_0}$. In this case the PLS estimate of C is

$$C_{PLS} = \sum_{j=1}^{l_0} \alpha_{l_0}^*(j)(\Sigma_X)^{l_0-j}(\Sigma_{XY}) = K_{l_0}\begin{bmatrix} \alpha_{l_0}^*(1)I_p \\ \alpha_{l_0}^*(1)I_p \\ \vdots \\ \alpha_{l_0}^*(l_0)I_p \end{bmatrix} = K_{l_0}\left(\alpha_{l_0}^* \otimes I_p\right) \quad (4\text{-}64)$$

where $K_{l_0} = K_l\big|_{l=l_0} \in \Re^{l\times(l_0 p)}$. Therefore, the PLS solution can also be expressed in similar way as the ordinary least squares solution. The number of columns l_0 in the reduced controllability matrix K_{l_0} can in principle be taken as the effective rank of the Krylov matrix K_l. In fact, it is proved in Di Ruscio (2000) that the column space of the weighting matrix W computed by the PLS algorithm and the column space of the reduced controllability matrix K_{l_0} coincide. In the next section we will explain how the coefficients of the characteristic polynomial are estimated.

4.5.4 Optimal Estimation of $\alpha_{l_0}^*$

In line with the least squares criteria, the parameter vector $\alpha_{l_0}^*$ is determined such that the prediction error matrix will have the minimum squared Frobenius norm (Di Ruscio, 2000). An estimate of the prediction error matrix can be obtained as follows:

$$Y - XC = Y - XK_{l_0}\left(\alpha_{l_0}^* \otimes I_p\right) \quad (4\text{-}65)$$

Defining

$$P_{l_0} = XK_{l_0} = \left[P_{l_0}^{(1)}\middle|P_{l_0}^{(2)}\middle|\cdots\middle|P_{l_0}^{(l_0)}\right] \quad (4\text{-}66)$$

Where $P_{l_0}^{(j)} \in \Re^{N\times p}$ is the jth submatrix for j = 1, …, l_0, we rewrite Eqn (4-65) as

$$Y - XC = Y - P_{l_0}\left(\alpha_{l_0}^* \otimes I_p\right) \quad (4\text{-}67)$$

We further define

$$D_{l_0} = \left[Vec\left(P_{l_0}^{(1)}\right) \quad \cdots \quad Vec\left(P_{l_0}^{(l_0)}\right)\right] \quad (4\text{-}68)$$

Where $Vec(\cdot)$ is the column vector operator of a matrix. For instance, $Vec(Y) \in \Re^{Np}$ is a column vector constructed from the data matrix Y by sequentially stacking each column of Y onto one another. Applying the column vector operator $Vec(\cdot)$ to Eqn (4-67) leads to

$$Vec(Y - XC) = Vec(Y) - D_{l_0}\alpha^*_{l_0} \tag{4-69}$$

As a consequence, minimizing the squared Frobenius norm of error matrix Y-XC in Eqn (4-67) is equivalent to minimizing the squared 2-norm of $Vec(Y) - D_{l_0}\alpha^*_{l_0}$ in the preceding equation. The least squares solution to this minimizing problem can be readily obtained as follows:

$$\alpha^*_{l_0} = \left(D^T_{l_0} D_{l_0}\right)^{-1} D^T_{l_0} Vec(Y) \tag{4-70}$$

Where the choice of l_o ensures the existence of $\left(D^T_{l_0} D_{l_0}\right)^{-1}$. The complete PLS algorithm is summarized in Table 4-6.

Table 4-6: Summary of Krylov controllability matrix (KCM) based PLS algorithm

1. Start with data matrix $X \in \Re^{N \times l}$, and $Y \in \Re^{N \times p}$
2. Calculate $\Sigma_X = \frac{X'X}{N-1}$; $\Sigma_{XY} = \frac{X'Y}{N-1}$
3. Choose an integer $l_0 < l$ and construct a reduced controllability matrix K_{l_0} as follows:
   $K_{l_0} = \left[(\Sigma_X)^{l_0-1} \Sigma_{XY} \quad (\Sigma_X)^{l_0-2} \Sigma_{XY} \quad \cdots \quad \Sigma_X \Sigma_{XY} \quad \Sigma_{XY} \right] \in \Re^{l \times l_0 p}$
4. Calculate $P_{l_0} = XK_{l_0} = \left[P^{(1)} \vdots P^{(2)} \vdots \cdots \vdots P^{(l_0)} \right] \in \Re^{N \times l_0 p}$; here P is a block matrix with elements $P^{(i)} \in \Re^{N \times p}$.
5. Using a vector operator which stack each column of $P^{(i)}$ onto one another, formulate a long matrix from P_{l_0} as follows:
   $D_{l_0} = \left[Vec(P^{(1)}) \vdots Vec(P^{(2)}) \vdots \cdots \vdots Vec(P^{(l_0)}) \right] \in \Re^{Np \times l_0}$
6. Calculate $\alpha_{l_0}^* = (D_{l_0}' D_{l_0})^{-1} D_{l_0}' Vec(Y) \in \Re^{l_0}$
7. Estimate PLS regression coefficient C as: $C_{PLS} = K_{l_0} \begin{bmatrix} \alpha_{l_0}^*(1) I_p \\ \alpha_{l_0}^*(1) I_p \\ \vdots \\ \alpha_{l_0}^*(l_0) I_p \end{bmatrix} \in \Re^{l \times p}$

4.5.5 Modified KCM based PLS theory and RPLS

It must be mentioned that although the KCM based PLS theory presented in Table 4-6 has a sound theoretical basis, it runs into some numerical problem. The problem is more pronounced with industrial data corrupted with noise. It was found that appropriate choice of l_0 is crucial for the numerical stability of the algorithm. Di Ruscio (2000) suggested l_0 to be equal to the rank of the complete controllability matrix K_l. However this choice of l_0 results in rank deficient D_{l_0}, which in turn makes the inverse of $(D_{l_0}^T D_{l_0})^{-1}$ non existent. A practical solution for the choice of l_0 is proposed as follows:

l_o must be chosen such that D_{l_0} is full column rank, hence inverse of $D'_{l_0} D_{l_0}$ exists. During implementation of this algorithm, we can select l_0 by successively adding one column to K_{l_0}, followed by construction of P_{l_0} and D_{l_0} and checking the rank of D_{l_0} matrix at each step. The process is continued until D_{l_0} becomes rank deficient. We select an l_o such that D_{l_0} is of full column rank, but D_{l_0+1} is rank deficient.

Even when using the above method to choose l_0, this algorithm may often result in ill conditioned K matrix due to round off errors when computing large powers of Σ_X. It turns out that often $l_0 << l$, which results in poor prediction. A higher value of l_0 is obtained with scaled data. Thus scaling improved the numerical stability of this algorithm to some extent.

To make the algorithm numerically robust, one can perform a QR decomposition of the Krylov matrix. QR decomposition (also known as the orthogonal-triangular decomposition) of a matrix expresses the matrix as the product of a real orthonormal or complex unitary matrix and an upper triangular matrix. This factorization is useful for both square and rectangular matrices. Performing QR decomposition on K_{l_0} results in:

$$K_{l_0} = Q_K R_K \tag{4-71}$$

Here $Q_K \in \Re^{l \times l_0 p}$ is an orthonormal and $R_K \in \Re^{l_0 p \times l_0 p}$ is an upper triangular matrix. Di Ruscio (2000) pointed out that the usual algorithm for computing the PLS weighting matrix W presented in the literature is equivalent to computing an orthogonal basis matrix (with orthonormal columns) for the column space of the Krylov matrix. This basis is equivalent to the orthogonal matrix Q_K obtained from QR decomposition of the controllability matrix. The QR decomposition is a numerically stable method for computing the column space of K_{l_0}. Moreover, the problem of computing an orthogonal basis for the controllability subspace may be better conditioned as compared to explicitly forming the controllability matrix. By using Q_k instead of K_{l_0} in all subsequent steps of the algorithm, the ill conditioning problem of the Krylov matrix can be avoided. However,

it still does not guarantee that the inverse of $D_{l_0}' D_{l_0}$ will exist if we use Q_k as an approximation of K_{l_0} in the calculations (steps 4, 5, and 6 in Table 4-6). Instead, we suggest performing a QR decomposition of the D_{l_0} matrix as follows:

$$D_{l_0} = Q_D R_D \tag{4-72}$$

Here $Q_D \in \Re^{Np \times l_0}$ is orthonormal, and $R_D \in \Re^{l_0 \times l_0}$ is an upper triangular matrix. Now, an estimate of C_{PLS} can be written as follows:

$$\alpha_{l_0} = \left(Q_D^T Q_D\right)^{-1} D_{l_0}' Vec(Y) \tag{4-73}$$

Finally, we calculate, C_{PLS} using Eqn (4-64). Note that, for univariate Y, $P_{l_0} = D_{l_0}$. Therefore, in that case, we perform QR decomposition on P_{l_0} instead of D_{l_0}. Table 4-7 summarizes a modified version of the KCM based PLS algorithm that has better numerical properties.

Table 4-7: Modified Krylov controllability matrix (KCM) based PLS algorithm

1. Start with mean centered and unit variance scaled data matrix $X \in \Re^{N \times l}$, and $Y \in \Re^{N \times p}$
2. Calculate $\Sigma_X = \frac{X'X}{N-1}$; $\Sigma_{XY} = \frac{X'Y}{N-1}$
3. Initialize $K = [\ ]$, $P = [\ ]$, and $D = [\ ]$
4. Set a = 1
5. Calculate

   $K = \left[(\Sigma_X)^{a-1} \Sigma_{XY} \quad K \right]$

   $P = [XK \quad P]$

   $D = [Vec(P) \quad D]$

   $r = rank(D)$
6. If $r = a$, set $a = a + 1$ and go to step 5, otherwise set $l_0 = a - 1$ and go to step 7
7. Formulate the reduced controllability matrix K_{l_0} as shown in step 3 in Table 4-6
8. Calculate $P_{l_0} = XQ_{l_0}$ and $D_{l_0} = Vec(P_{l_0})$
9. Perform QR decomposition of D_{l_0} such that $D_{l_0} = Q_D R_D$
10. Calculate $\alpha_{l_0} = (Q_D^T Q_D)^{-1} D_{l_0}' Vec(Y) \in \Re^{l_0}$
11. Estimate PLS regression coefficient C as follows: $C_{PLS} = K_{l_0} \begin{bmatrix} \alpha_{l_0}(1) I_p \\ \alpha_{l_0}(1) I_p \\ \vdots \\ \alpha_{l_0}(l_0) I_p \end{bmatrix} \in \Re^{l \times p}$

Using the modified algorithm shown in Table 4-7, a block wise recursive PLS algorithm may be developed. Implementation of the block wise RPLS algorithm may be done in a similar manner as the NIPALS based RPLS method. The complete RPLS algorithm for online implementation can be summarized in the following steps:

a) Start with an initial block of data X and Y, having N samples of l inputs and p outputs
b) Mean center the data and scale to unit variance
c) Calculate an initial estimate of C_{PLS} using the algorithm presented in Table 4-7
d) When new data are available, select a window size w; select the data block to be used for updating; denote the new data as X_1 and Y_1
e) Estimate updated covariance matrices $\Sigma_{X,new}$ and $\Sigma_{XY,new}$ using scaling parameters obtained from the previous block
f) Follow steps 3 to 11 in Table 4-7 and get an updated estimate of $C_{PLS,new}$
g) Repeat with new data blocks

It is important to note some characteristic features of this RPLS method as compared to the NIPALS and the kernel based methods presented earlier. The difference between this algorithm and the existing RPLS algorithm can be summarized as follows:

1. The PLS calculation is performed using the Krylov controllability theory, which is quite different from the NIPALS and the kernel algorithm.
2. This algorithm is non-iterative. Both NIPALS and kernel algorithms for PLS are iterative for multivariate Y
3. Even for univariate Y, the PLS solution is obtained directly from the Krylov controllability matrix and the coefficients of the characteristic polynomial. In the NIPALS and the kernel algorithm, the PLS vectors are calculated one at a time, in an iterative manner.
4. This algorithm completely avoids any deflation and singular value decomposition. Both NIPALS and kernel algorithms involve these operations.

Since no deflation, SVD, and iteration are involved, it is expected that this algorithm will result in much faster implementation compared to the other existing methods.

4.6 Grade Transition, Data Quality & Artifacts of MI Readings

In most polymer plants, it is common to produce many product grades using a single reactor. The LDPE/EVA reactor that is studied in this project produces 26 grades of EVA copolymer on a regular basis. Melt Index for these various products range from as low as 0.3 to as high as 2000. To accomplish this change in MI, operating conditions such as reactor pressure, temperature, initiator composition and type, propylene concentration etc. are varied over a wide range. The change from one MI grade to another is called grade transition. Duration of a grade transition usually depends on the difference in the final properties of the two products.

A major obstacle encountered in this work was obtaining reliable grade transition data. The MI estimated by the online analyzer is stored in the historian as a 4-20mA (milliampere) signal. During storage of the data, all MI values are scaled to a two digit number ranging between 0 - 99.99. This is a pre-defined setting of the data storage facility at the plant, which was designed at the time the MI analyzer was installed. For those grade changes where the true value of MI falls within this range, there is no discrepancy between the stored value and that obtained from the analyzer. If true MI > 99.99 then the stored value does not match the real value of MI.

However, even for MI < 99.99, often the MI value estimated by the analyzer is not reliable during transition. This is due to the fact that the operators sometime manually change the die in the online MI unit. The equipment is set on standby mode for this duration. The stored MI value is kept constant at its previous level using a zero order hold or stored as zero during the standby mode. The data collected during such grade transition thus shows many constant segments, missing values or false spikes, which should be discarded from the training data. Figure 4-1 shows a trend plot of Melt Index during an EVA grade transition. This is one of the most common product changes that take place at the plant. The target change in melt index is from 6 to 4. The duration of this change was 1.5 hours. Since the magnitude of the actual MI for both grades falls within the range (0 - 99.99), these values are stored in the historian without any scaling

distortion. However, even then, the data during transition was of poor quality, and contaminated with many false values.

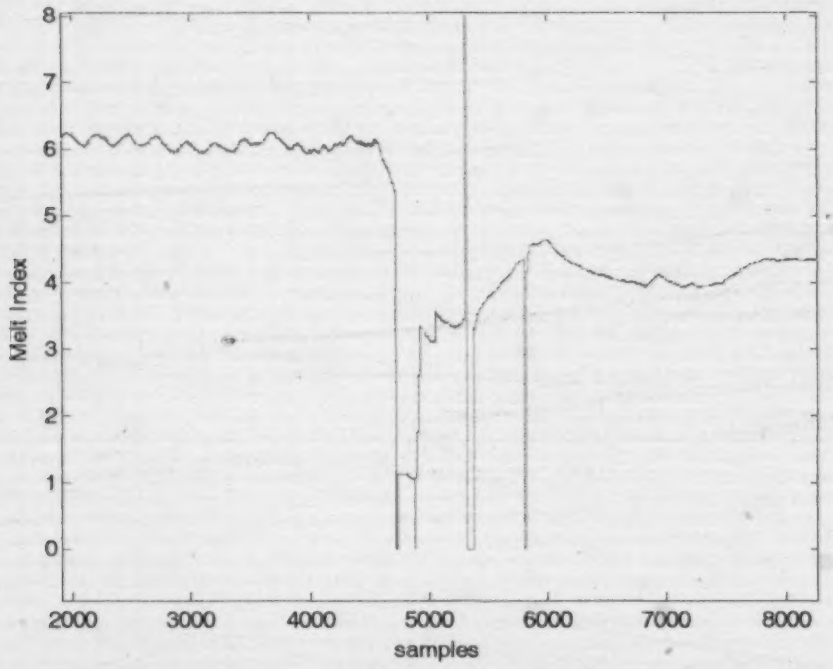


Figure 4-1: Melt Index Change during a Grade Transition from 6 to 4 MI. Note the time when MI is constant due to the standby mode.

The problem is more pronounced during grade change where melt index undergoes an order of magnitude change (such as from MI < 100 to MI > 100). Since online melt index values are always stored in the historian as number between 0 and 100, this creates confusion in cases where true melt index is higher than 100. Figure 4-2 shows MI values during a grade change operation where the MI changes from 6 to 445. Due to inaccurate scaling, stored MI values for the final grade shows a value around 44.5 whereas the actual value is near 445. In addition, the values during the transition are erratic due to this data storage technique. This makes the interpretation of the data in the transient region extremely difficult. This problem persisted throughout the project in spite of repeated

recommendations made to the plant personnel to rectify the scaling issue. As a result, the data during the transition had to be collected carefully. Only transitions with small changes in MI were considered for this work. Significant effort was given to make sure that the data was reliable. As a result, only a limited amount of good quality data was available to model grade transition. In should be mentioned that very recently a melt index soft sensor based on data from the production extruder developed by Alleyne, 2006 has been implemented at the plant. Melt index obtained from this softsensor model are stored accurately in the historian without any scaling problem. However, the work presented in this thesis was completed before implementation of this new softsensor. It may be mentioned in this regard that the PLS based soft sensor developed in this work is based on reactor operating conditions; hence it shows the MI values approximately 10-12 minutes before the extruder based soft sensor.

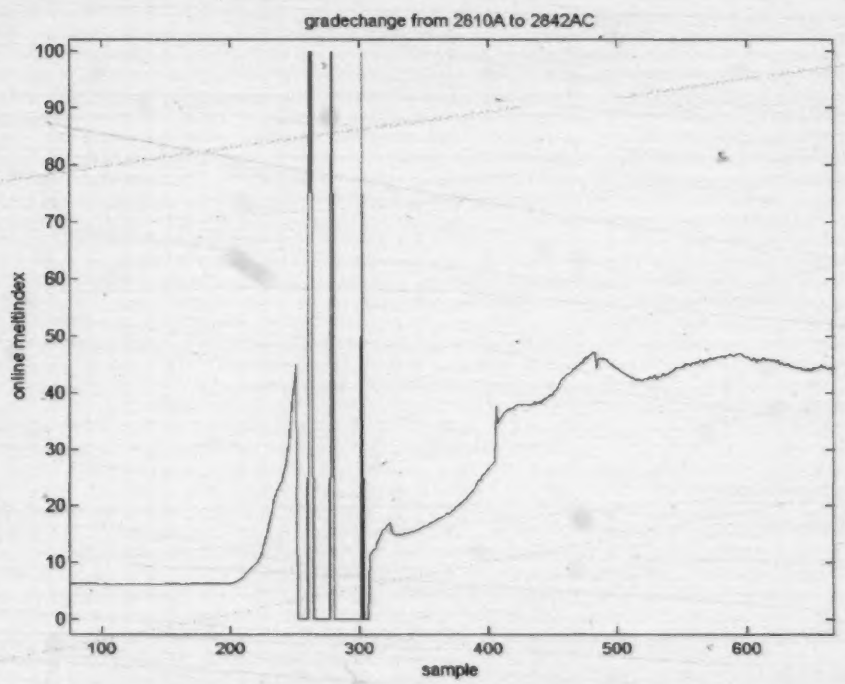


Figure 4-2: Online Melt Index during a grade change from 6 to 445 MI

4.7 Modeling Grade Transition

In chapter 3, we showed a simple bias updating scheme for implementation of the PLS based melt index soft sensor. The method was shown to be adequate to rectify small operational changes during production of the same polymer grade. However, during product changes, the process often moves to a very different operating condition. A model built from steady state data will not be applicable during transient period. In order to examine how a static model performs during grade transition, a dataset was collected containing 11 inputs and one output variable. Inputs for the model included fresh feed flow rates (ethylene, VA, and propylene), reactor zone temperatures and reactor pressure. Sampling time for the data collection was 5 seconds. The data included measurement of all variables during steady production of the initial and the final grade before and after transition, as well as transient data during the transition. All preprocessing steps described in Chapter 3 such as checking for data compression, filtering, outlier detection were followed.

Figure 4-3 shows a plot of property changes during an EVA grade transition that included relatively large change in MI and VA. The desired change in MI was from 150 to 30 and the comonomer (vinyl acetate) content was changed from 23% to 19% (by mass). Figure 4-4 and 4-5 show the changes in the manipulated variables that were made in order accomplish these property changes.

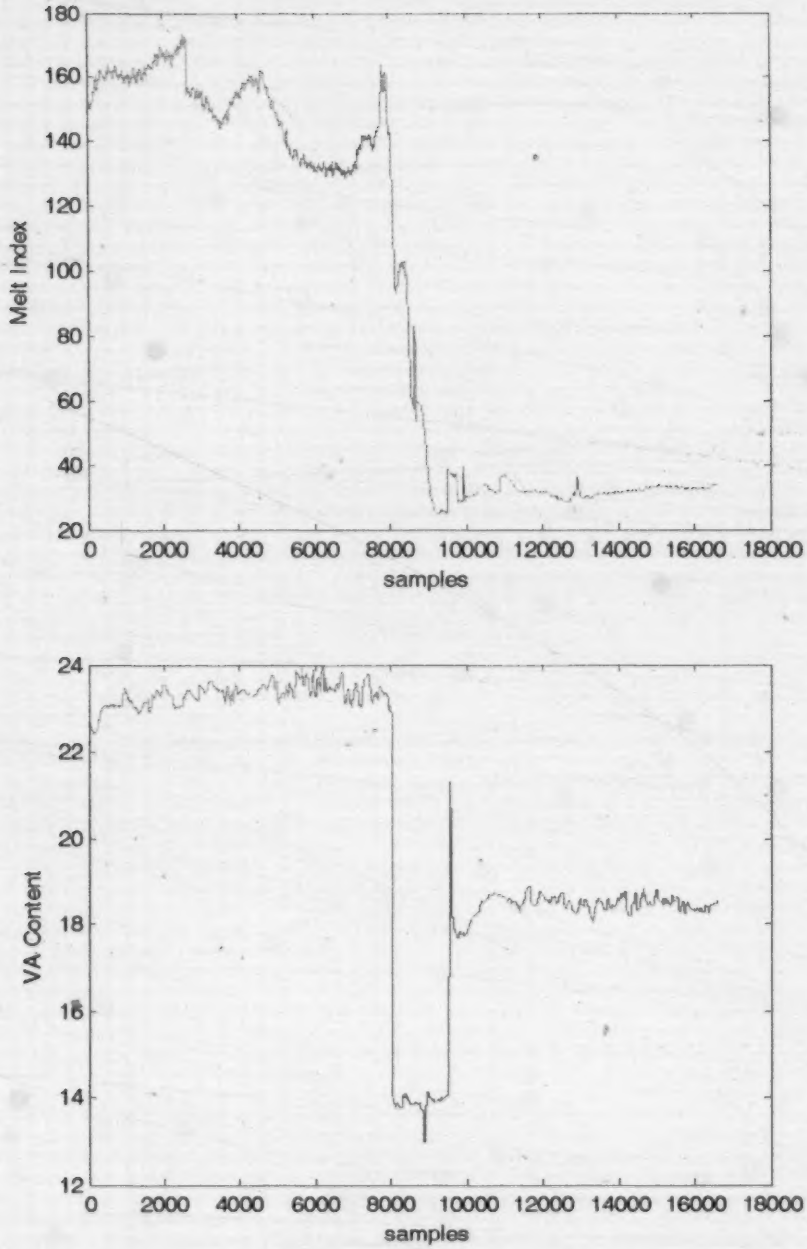


Figure 4-3: Change in product properties during grade transition; Top: change in melt index; Bottom: Change in Vinyl Acetate content

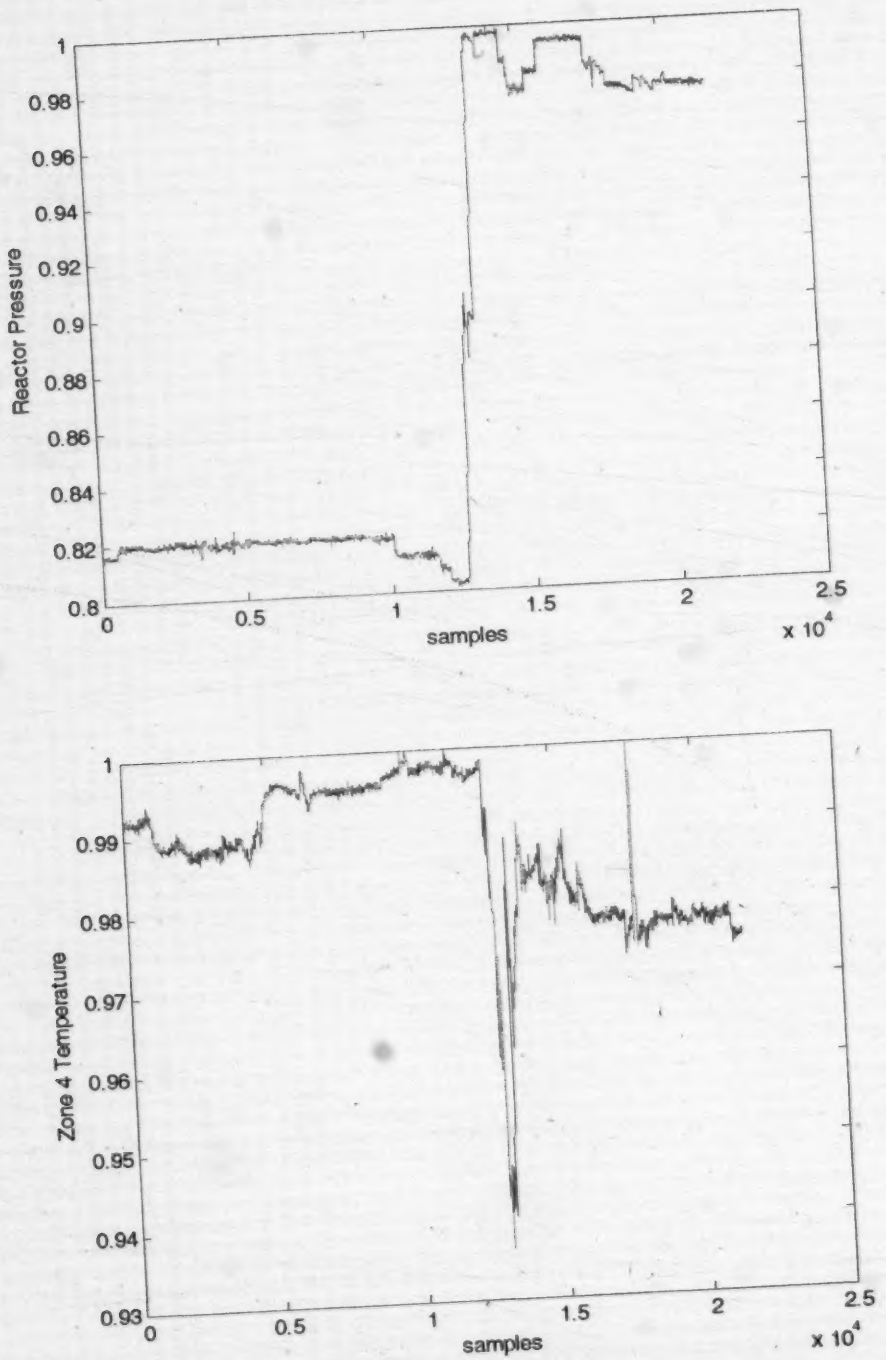


Figure 4-4: Changes in scaled reactor pressure (Top) and reactor temperature (Bottom) during grade transition

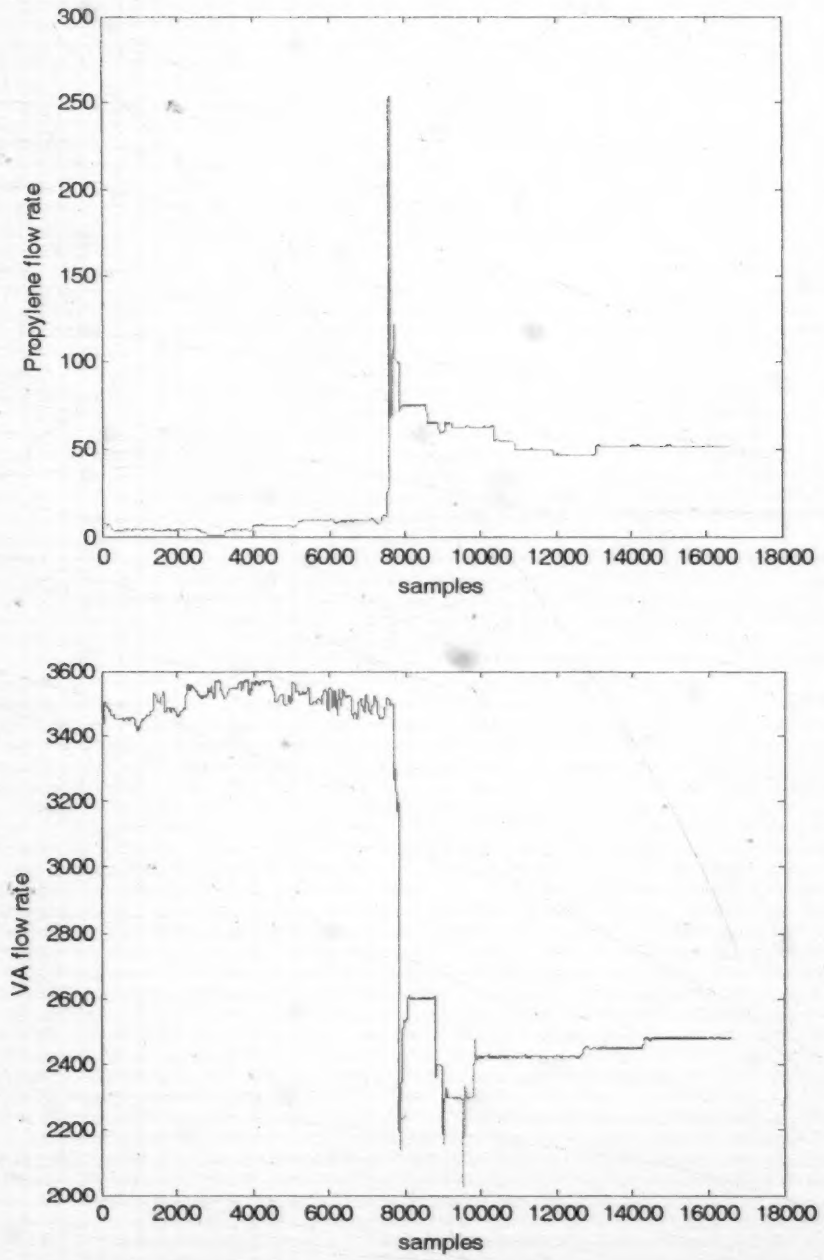


Figure 4-5: Changes in propylene (Top) and VA mass flow rate (Bottom) during grade transition

From approximately sample # 5000, reactor pressure and flow rate of propylene started to change. Subsequently the comonomer flow rate and the rate of initiator addition into the reactor were manipulated resulting in the desired changes in the melt index and the comonomer content. Duration of the transition was about 6 hours. A PLS model was built using the first 5000 samples collected from the initial grade having a melt index of 150. This model was then applied to predict MI during and after the transition. Figure 4-6 shows the prediction from the model, which completely failed to predict the transition. When the model quality deteriorates significantly, bias update using spot values of the melt index is not a practical solution. It is evident that a new model should be built or the old model needs to be updated using data taken from the transition, in order to include the process changes. A similar result is obtained when a constant PLS model was used to predict a relatively smaller grade transition (6 MI to 4 MI). This was the motivation for using the adaptive PLS modeling techniques.

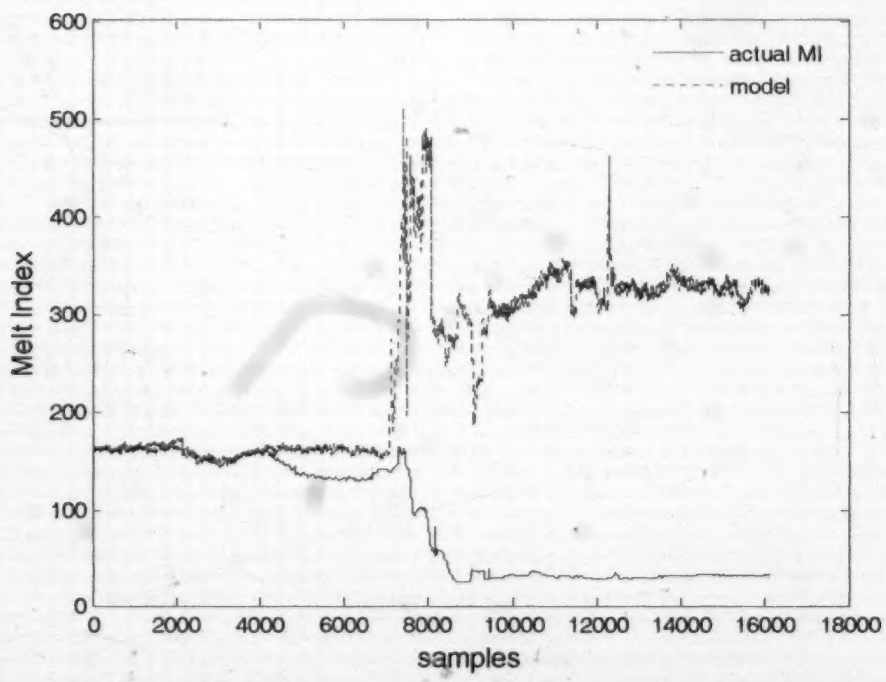


Figure 4-6: Prediction from a constant PLS model during and after grade transition

4.8 Implementation of the RPLS Algorithms

In this section we present the results obtained from implementation of the three RPLS methods discussed in the earlier sections. For ease of comparison, the same set of data was used in all cases. The data set includes an EVA grade transition operation where MI changes from 6 to 4. Comonomer content of the products was changed from 28% to 26% (by mass). Duration of the grade transition was approximately three hours. The initial and the target grades of polymers were fairly similar in properties. A transition with a relatively small change in properties was selected for this analysis for reasons explained in section 4.6. In order to accomplish the property changes (in terms of MI and VA content), mass flow rate of the comonomer and the reactor pressure were used as the manipulated variables. Initiator flow rates into the reactor were also changed which resulted in changes in the reactor temperatures. Propylene was not used in any of these two products; hence flow rate of propylene was not included in the data. The final data set contained 12 inputs and one output variable.

4.8.1 NIPALS based RPLS Method: Results

The RPLS algorithm based on the NIPALS method was first applied on the test data set. The data consisted of 1350 samples of 12 inputs and 1 output variable. Sampling time for data collection was one minute. A moving window of fixed length was used for the adaptive implementation. An initial PLS model was built using the first 120 samples (corresponding to 2 hours). A window length of 10 (10 samples) was selected to collect new blocks of data and subsequently update the model at the end of each window. The initial model denoted as model-1, was used for prediction for the next 10 samples (sample#121-130). Next, model-1 was updated by combining it with the samples collected in the past ten minutes. This updated model (denoted as model-2) was used for prediction for the next window of samples (sample#131-140). Again, model updating was performed by combining model-2 with sample #131-140 and was used for prediction during the next 10 samples, and this process was repeated.

A question that is not answered definitively in any data based modeling with non-stationary data is how the data should be scaled. Neither Helland *et al.* (1991) nor Qin

(1998) explain any adaptive scaling technique in their algorithm. Two suggestions on data scaling found in their work are as follows:

Scale the first block of the data to zero mean and unit variance or using any other predefined scaling constants. These scaling parameters can be used to scale all subsequent data blocks. But for a time varying process, the mean and variance may change over time. Therefore, scaling factors obtained from the initial block of data will not make subsequent data blocks zero mean and unit variance. However, the RPLS algorithm is still expected to work with data whose variance may change over time.

The other option is not to mean center or scale the data. However, if the mean of each variable in the data matrices is not zero, the input-output relationship has to be modified with the following general linear relationship:

$$y_i = C^T x_i + d = \begin{bmatrix} C^T & d \end{bmatrix} \begin{bmatrix} x_i^T & 1 \end{bmatrix}^T \tag{4-74}$$

Where x_i and y_i represent the ith row of X and Y respectively. $d \in \Re^p$ is a vector of intercepts for the general linear model. Therefore to model data with non zero mean, the RPLS algorithm can be applied on the data pair $\{ \begin{bmatrix} X & 1 \end{bmatrix} \quad Y \}$. Here $1 \in \Re^N$ is a vector whose elements are all one.

Based on these different approaches for data scaling, we considered the following three cases:

a. data without any scaling
b. data with constant scaling parameters
c. data with variable scaling parameters

In each case the performance of the model was measured by estimating the mean square error of prediction and the R2 value. These performance indices are defined as follows:

$$MSE = \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N} \tag{4-75}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{4-76}$$

Here y_i and $\hat{y}_i$ represents the actual and the predicted output at observation number i. $\bar{y}$ represents the mean value of y. Figure 4-7 shows the results obtained in the first case. In this case the data was not scaled, but a unity column was appended to the X block to handle non-zero mean. The length of the initial training block of the data was 120. A moving window of 10 samples was used throughout this application. The solid line in Figure 4-7 shows the actual MI values obtained from the lab. The dotted line shows the result when model-1 was used for predicting MI for the entire range of data without performing any updating. The R^2 value was found to be -2.96, indicating very poor performance of the model. The mean squared error for prediction was 4.03. The constant PLS model performed poorly in predicting the product change. The dashed line shows the predicted values obtained with the RPLS algorithm. In this case, the R^2 value was found to be 0.92. The mean squared error (MSE) for prediction was 0.0814. It is clear that updating the model with a new block of data improves the predictive capability of the PLS models. All the PLS and RPLS algorithms were written as MATLAB codes and were run on a Pentium-4 3.2 GHz desktop processor. The time elapsed for running the RPLS algorithm on the entire data set (having 1350 samples of 12 inputs and one output) was 2.6 seconds.

In the 2nd case, the initial block of data was scaled to zero mean and unit variance before PLS model building. The mean and variance of this initial block was used to mean center and scale all subsequent blocks of data at each updating step. Figure 4-8 shows the results from the RPLS algorithm using the scaled data. The dotted line shows the results from the constant PLS model built with auto scaled data. The dashed line shows the results from the RPLS algorithm. Table 4-8 shows the calculated R^2 value and the mean squared error (MSE) for the constant PLS model and the RPLS models. The prediction quality between the first two cases was comparable as seen by visual inspection, R^2 value and the MSE values.

In the 3rd case study, the initial block of data is auto scaled using the mean and variance calculated from the same block. The mean and the variance from this block of data were used as the scaling parameters for the second data block. Each subsequent block of data was mean centered and scaled by scaling parameters calculated from the previous block of data. Figure 4-9 shows the results obtained using this scaling technique. Again, the solid, dotted and the dashed lines represent the actual MI, predictions from a constant PLS model, and RPLS algorithm respectively. The RPLS algorithm in this case results in much improved prediction when compared to the first two cases. Table 4-8 shows that the R^2 value and the MSE are significantly improved using variable scaling parameters.

From the three cases, it may be concluded that data scaling has significant impact on model performance. Adaptive scaling seems to produce the best result. This method is quite feasible for practical implementation. It was found that when the basic NIPALS algorithm for PLS is used for modeling, scaling gives very reasonable results based on the amount of variance explained by each LV. This makes the selection of number of latent variables more logical.

In this regard, it is worth mentioning that in the RPLS algorithm, decision on how many latent variables (or PLS factors) to retain in the model was made based on the Frobenius norm of deflated X. PLS factors were calculated until $\|E\| \approx 0$. This resulted in including all the LV's for the present data set. The RPLS algorithm includes the maximum possible number of PLS factors which is given by the rank of X (r = Rank(X)). However, when dealing with noisy data, rank determination using commercial software such as MATLAB is not a foolproof method and often gives erroneous results. Even when r factors are used in the PLS model, the PLS regression approaches the ordinary least squares solution. Qin (1998) has mentioned this issue in his work, which was also verified by Wang *et al.*, (2003). Qin (1998) suggested that the number of factors for model updating should make the input residual E close to zero, while the number of factors for prediction should be determined by cross validation. This is a confusing statement as the updated model which is developed by combining recent and past data is intended to be used for predicting future output values. The model should be used for its

intended applications without any modification. A rather disturbing fact in Qin's (1998) work was that he only showed the results of the RPLS algorithms using the training data. The updated models were not used for making prediction but rather refitting the training data. The results shown in this section use the input residuals as the only criteria for selecting the LVs at each updating step. The updated model was used without modification for predicting the future MI values during the subsequent window. Any modification to the initial model would make theorem 4-1 invalid. Therefore, we feel that the results shown here represent the most appropriate way of applying the method for predicting future output values in an adaptive framework.

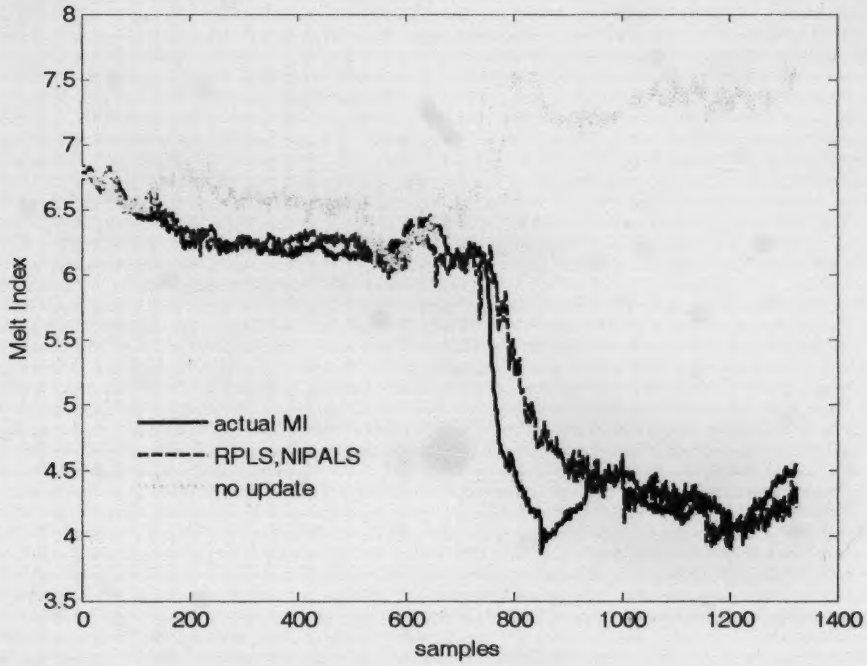


Figure 4-7: Results from the NIPALS based RPLS algorithm on unscaled data

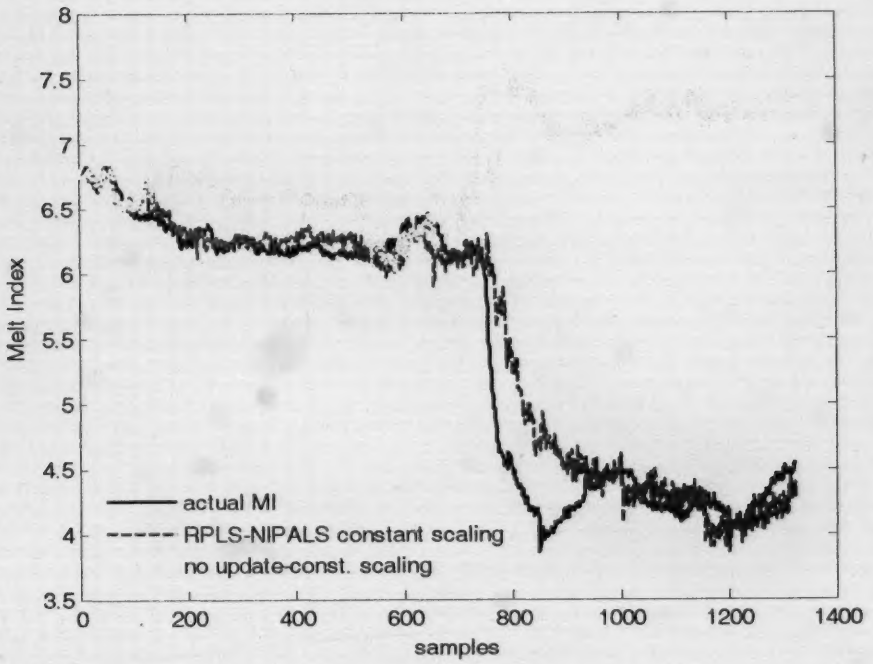


Figure 4-8: Results from the NIPALS based RPLS algorithm on scaled data with constant scaling parameters

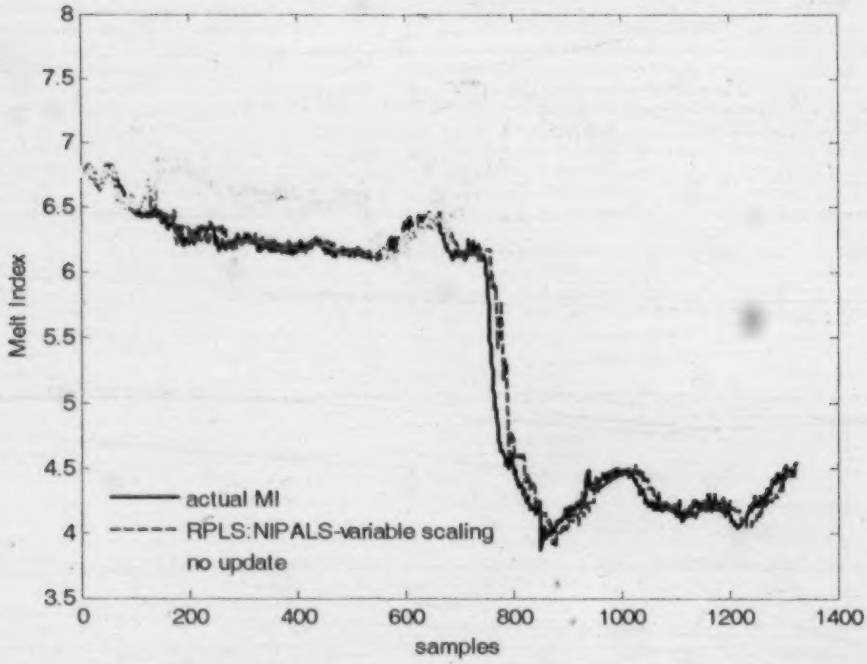


Figure 4-9: Results from the NIPALS based RPLS algorithm on scaled data with variable scaling parameters

Table 4-8: Summary of NIPALS based RPLS Model Performance

Scaling type	Modeling method	R^2 Value	MSE	CPU Time (seconds)
Case 1	PLS	-2.96	4.0295	-
	RPLS	0.9199	0.0814	2.50
Case 2	PLS	-2.70	3.7601	-
	RPLS	0.9190	0.0823	2.60
Case 3	PLS	-2.70	3.7601	-
	RPLS	0.9791	0.0243	2.04

4.8.2 Kernel based RPLS Method: Results

The original kernel based RPLS algorithm was developed using a sample wise updating scheme. The algorithm starts with an initial model which is trained using a finite number of samples. This model is used for predicting future outputs. As new measurements of the output variables become available (from lab analysis etc.), the model is updated at each of these output samples.

Regarding the issue of mean centering and scaling the variables for PLS model, Dayal & MacGregor (1997) suggested similar approaches as presented for the NILAPS based RPLS method in the previous section. For a time varying process, they suggested an adaptive calculation of the standard deviation of each process variable. At each sampling interval, the updated standard deviation can be computed from the updated $(X^T X)_t$ matrix. In this way, no extra memory or computational effort is required to store the standard deviation of the variables for each updating step.

The same data set that was used in the previous section was used for implementing the kernel based RPLS algorithm. The following two cases were considered based on the frequency of updating:

- RPLS with sample wise updating
- RPLS with block wise updating

In each case, a unity column was added with the X block and the data was used without scaling. The first 120 samples were used for building an initial PLS model. The number of PLS factors to be retained in the model was determined using cross validation. This resulted in retaining 8 factors in the initial model. In the first case study, the model was updated at each subsequent output sample. This resulted in almost perfect agreement between the actual and the predicted values of MI. Figure 4-10 shows the results obtained from the sample wise updating. The prediction obtained from the constant PLS model built from the initial data set is also shown in the same plot. The R^2 value estimated from the constant PLS model and the RPLS model were -2.83 and 0.9959 respectively. MSE for prediction was estimated as 3.898 and 0.004 from PLS and RPLS method. The model's ability to predict well is quite expected as the updated model is used only to

make a one step ahead prediction. While this results in excellent performance in terms of prediction, this is not a very practical solution for online implementation. This will be very computationally intensive for large data matrices.

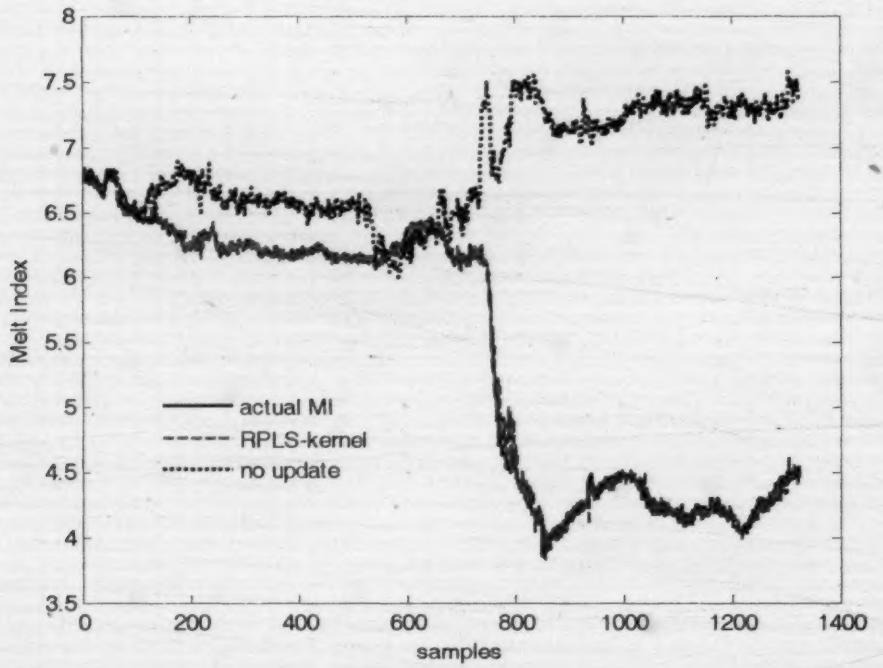


Figure 4-10: Results from the kernel based RPLS algorithm using sample wise updating

In the 2nd case study, we selected a moving window of 10 samples. The initial model which was built using the first 120 samples was used to predict MI during samples 121-130. At sample # 130, the covariance matrices were updated using all the data collected during the window. An updated PLS model was built using the new covariance matrices and was used for predicting during the next window of samples. Figure 4-11 shows the results obtained using this approach. Table 4-9 shows a summary of the performance of the RPLS model for the two types of updating schemes. Using the moving window approach, the quality of the model becomes slightly worse than the first case as seen from visual inspection and R^2 values between the actual and the predicted values. Mean

squared error for prediction is comparable in both cases. However, the block wise updating is much faster than the sample wise updating scheme. Hence, this is a more reasonable method for online implementation as the computational effort for updating the model at each sample particularly if the data matrix contains many variables may become considerably large.

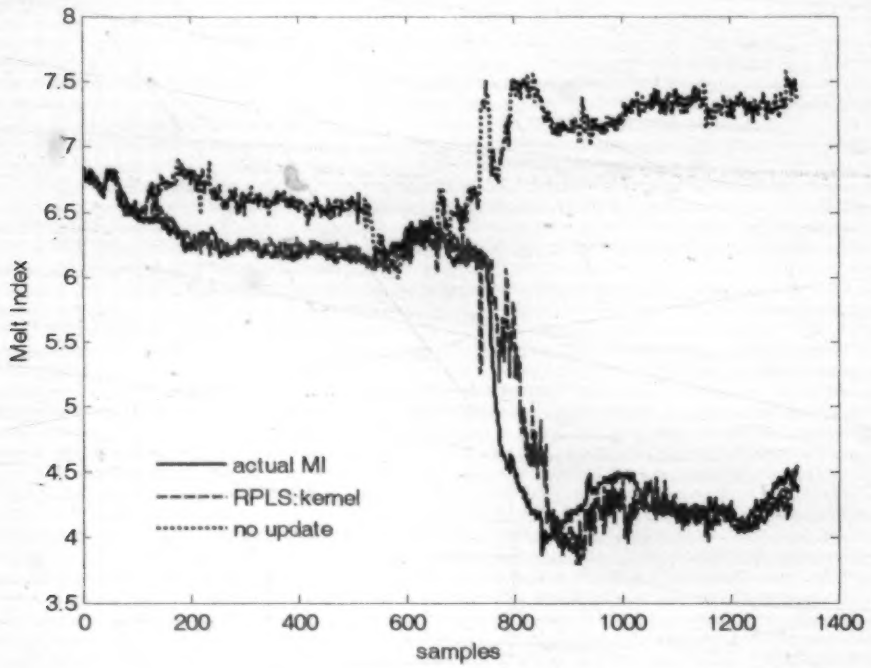


Figure 4-11: Results from the kernel based RPLS algorithm using block wise updating

Table 4-9: Summary of Kernel based RPLS Model Performance

Frequency of Updating	Modeling method	R^2 Value	MSE	CPU Time (seconds)
Sample wise update	PLS	-2.83	3.898	-
	RPLS	0.9959	0.004	2.00
Block wise update	PLS	-2.83	3.898	-
	RPLS	0.9413	0.0596	1.42

4.8.3 KCM based RPLS Method: Results

In this section we present the results obtained from application of the controllability based RPLS algorithm on the same data matrix. Similar to the other two methods, the first 120 samples were used to build an initial model using the KCM based PLS algorithm. A moving window of 10 samples was used for recursive updating of the model as outlined in section 4.5.5. The Krylov matrix K_{t_0} was constructed using one column at a time approach. This resulted in retaining 5 columns in K, which is equivalent to the number of PLS factors in the previous RPLS methods. Figure 4-12 shows the results from the RPLS algorithm. The RPLS algorithm can successfully predict the data during transition compared to the case when the initial model is used without performing any updating. The R^2 value was -0.68 and 0.89 using the constant PLS and the RPLS method respectively. MSE for predicted MI was 1.24 and 0.134 for the PLS and RPLS methods respectively. It should be noted that when the KCM based RPLS method was used, the quality of model prediction was poorer than the NIPALS based and the kernel based RPLS algorithms. This assessment of quality is mainly in terms of the lower R^2 value and the higher MSE resulting from the KCM based approach. However, it should be noted that both NIPALS based and kernel based methods contained more factors. Moreover, the KCM based method requires the least time for computation. It is even faster than the kernel based RPLS algorithm.

In applications where more emphasis is given on better prediction than the speed of the algorithm, a shorter window length may be used. In a second case study, a shorter window length of 5 samples was used for updating the RPLS model. This resulted in some improvement in prediction as evident from a lower MSE and a slightly higher R^2 value. Time required to run the algorithm was slightly greater. Figure 4-13 shows the results obtained with the shorter window length. Table 4-10 summarizes the performance of the RPLS models using different moving window lengths.

As a final comment on model performance, it should be mentioned that the time required for running each algorithm may seem quite negligible for a high performance desktop computer. However, actual time required for online implementation of the algorithms in

industrial settings will be much higher due to hardware limitations and data transmission delays. The Krylov based RPLS algorithm will be a more appropriate alternative in this regard.

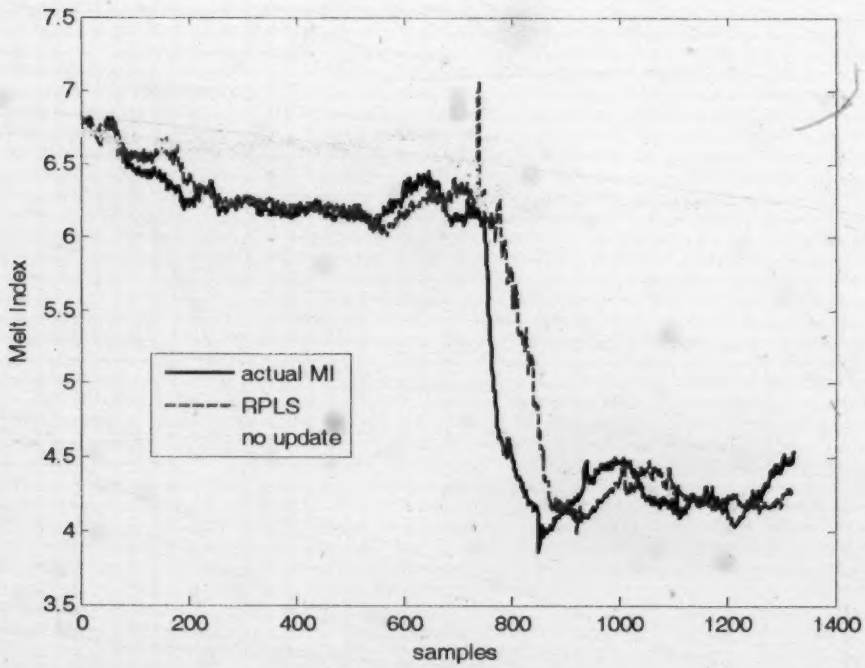


Figure 4-12: Results from the KCM based RPLS algorithm using a moving window length of 10

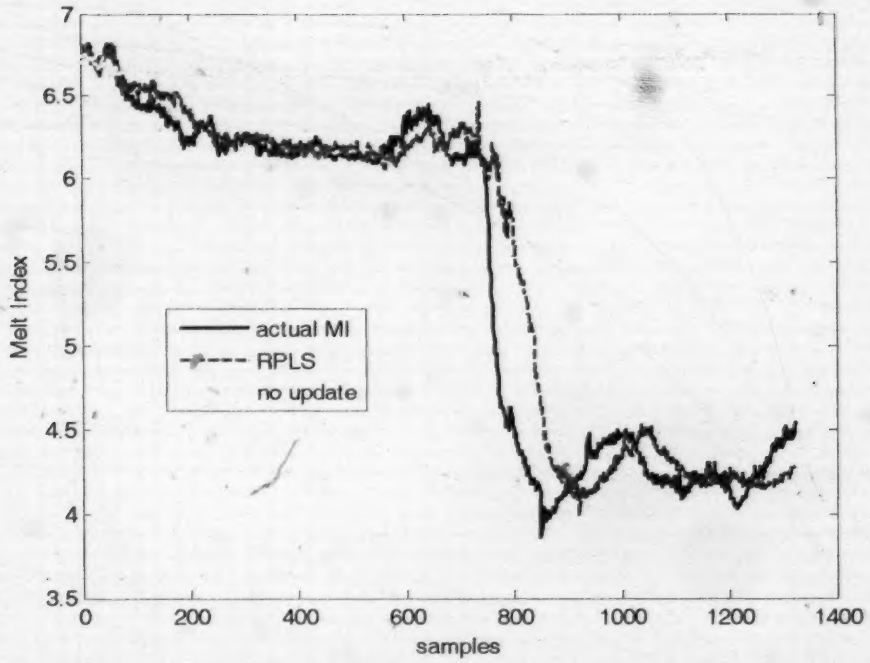


Figure 4-13: Results from KCM based RPLS algorithm using a moving window length of 5

Table 4-10: Summary of Controllability Matrix based RPLS Model Performance

Length of moving window	Modeling method	R^2 Value	MSE	CPU Time (seconds)
w = 10	PLS	-0.6896	1.2385	-
	RPLS	0.8942	0.1340	1.03
w = 5	PLS	-1.559	2.6020	-
	RPLS	0.9053	0.0963	1.16

4.9 Summary

In this chapter we presented a detailed theoretical review of all the existing recursive partial least squares algorithms used to model time varying systems. Some minor modifications were suggested for both the NIPALS and kernel based RPLS algorithms to make them more suitable for practical implementation. In addition, a new interpretation of PLS using a controllability matrix based approach has been introduced. A number of modifications were suggested to make the original KCM based PLS algorithm numerically robust. A block wise RPLS method was suggested based on the modified version of the controllability based PLS. Three recursive algorithms were applied to model melt index of EVA polymer during a small grade transition operation. All the RPLS methods were successful to predict the transient data. RPLS based on NIPALS algorithm shows good performance when applied with variable scaling. This method is simple to implement online and can be considered for industrial use. Although kernel algorithm was shown to be faster than the NIPALS algorithm for the basic PLS model building, for cases where only one predicted variable is involved, this results in negligible improvement in terms of computational efficiency since no iteration is necessary for computing the PLS vectors. Moreover, updating the model at each new sample can be quite computationally expensive and unfeasible for online applications. RPLS based on KCM resulted in a very fast approach with the least number of PLS factors. This method is expected to have much industrial importance in modeling time varying linear systems.

It was observed from an extensive literature review on PLS and RPLS algorithms that the developed theories are seldom validated with real data collected from continuous processes. An important contribution from this chapter is that it reduces the gap between theory and its practical implementation. The work presented in this chapter is the first attempt to apply the RPLS theories to model a complex industrial polymerization process.

Chapter 5

PCA Based Reactor Monitoring

5.1 Introduction

High pressure LDPE reactors operate under supercritical conditions. Normal operating pressure in these reactors can reach 3500 atm. Therefore, it is extremely important to monitor their performance in order to prevent any abnormal situations. An undesirable and catastrophic phenomenon that happens sporadically in high pressure polyethylene reactors is known as ethylene decomposition. This is a highly explosive exothermic reaction that leads to significant damage to the process and cause significant economic loss. Therefore, any monitoring strategy capable of detecting decomposition in high pressure ethylene reactors will be of great industrial importance.

In this chapter, we present a monitoring method for an autoclave reactor using an overall reactor energy balance. An appropriately formulated energy balance shows the sources of heat generation and means of removal of heat from the reactor and should close within reasonable limits when the process is running under normal conditions. If excess heat is generated inside the reactor, and the heat is not removed as soon as it is being generated, this may eventually lead to decomposition. In such cases, the energy balance closure error would increase and this would give an indication of an impending decomposition. However, for many practical systems encountered in chemical industries, all the necessary information required to formulate the balance equation may not be available.

Part of the material presented in this chapter has been submitted to: Sharmin R., Shah, S.L., Sundararaj, U., A PCA based Fault Detection Scheme for an Industrial High Pressure Polyethylene Reactor, *Macromolecular Reaction Engineering*, 2007.

For example, for the LDPE reactor, flow rates and the chemical identity of all inlet and outlet streams, heat capacity of reactants and products at supercritical conditions, the extent of reaction etc. are not readily available from routine operating data. For such cases, solving the balance equation with unknown parameters may be considered as a model identification or parameter estimation exercise. Multivariate statistical techniques such as principal component analysis (PCA) can be used as a model identification tool to solve this problem.

Principal component analysis was introduced by Pearson (1901) and later extended by Hotelling (1947). Kresta *et al.* (1991) and MacGregor *et al.* (1991) developed the method further and explained the basic methodology to use PCA as a monitoring tool for continuous processes. Formulated as a multivariate process control task, these applications use PCA to extract a few independent components from a set of correlated process data and use the components to monitor the process operation. The application and use of PCA in process monitoring has been demonstrated by many researchers through simulation (Kasper and Ray, 1992; Ku and Georgakis, 1995; Raich and Cinar, 1995; Luo *et al.*, 1999 etc.) and industrial case studies (Wise *et al.* 1991; MacGregor and Kourti, 1995; Li *et al.*, 2000 etc.). Among few published papers on the application of statistical process monitoring to polymerization processes, some examples include the use of PCA for fault detection and isolation in an LDPE reactor (Kumar *et al*, 2003), monitoring an industrial polymerization reactor (Piovoso, 1992), and product quality design for a high pressure polymerization process using statistical techniques (Moteki & Arai, 1998). Although primarily used as a compression tool, in recent days, PCA is gaining importance as a model identification technique (Ku *et al.*, 1995; Cao and Gertler, 2004, Gertler and Cao, 2004; Narasimhan and Shah, 2007 etc.). In this project, we used PCA to identify the energy balance relationship for an autoclave reactor, and then monitored the closure error from the balance equation to detect abnormal behavior. The study was conducted using data collected from a number of decomposition incidents that took place in an industrial LDPE/EVA autoclave reactor located at AT Plastics Inc., Edmonton, Canada.

The remaining sections of this chapter are arranged as follows: a short description of decomposition in high pressure polyethylene reactors is given; this is followed by an overview of PCA and its application in fault detection and model identification; next a description of the process, data collection, and formulation of an overall reactor energy balance are explained; finally, results from a number of case studies are presented.

5.2 Decomposition in High Pressure LDPE Rectors

LDPE and EVA are commercially produced by high-pressure free radical polymerization of ethylene. The reaction is carried out at pressures ranging from 1000-3500 atm and temperatures ranging from 140-330^{0}C in tubular or stirred autoclave reactors in presence of free radical initiators like peroxides, azo compounds or oxygen. The principal steps involved in both processes include initiation, propagation and termination. Propylene is sometimes used as a chain transfer agent for the termination steps. Polymerization of ethylene is a highly exothermic process with a standard heat of reaction of 807 Cal/g of ethylene polymerized (Mannan, 2006). The overall reactions for the formation of LDPE and EVA can be written as follows:

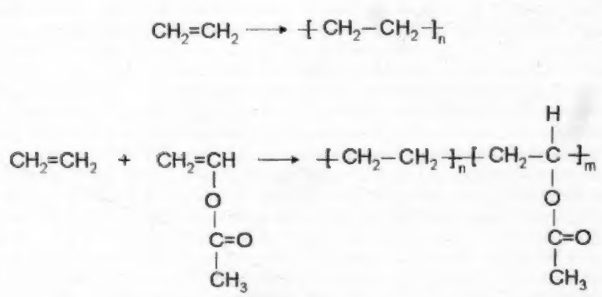


Figure 5-1: Polymerization of ethylene into polyethylene (top), and polymerization of ethylene and vinyl acetate to form ethylene vinyl acetate copolymer (bottom)

Under normal operating conditions in the polymerization reactors ethylene can undergo two further exothermic reactions where ethylene is decomposed into carbon, hydrogen and methane.

$$CH_2{=}CH_2 \rightarrow 2C + 2H_2$$

$$CH_2{=}CH_2 \rightarrow CH_4 + C$$

Above 350^0C, the decomposition reaction becomes faster than the polymerization reaction. Heat released and the consequent temperature rise leads to very rapid and complete decomposition unless heat can be removed as fast as it is being generated. The final temperature of the products of a complete adiabatic decomposition is 1200-1400^0C higher than the starting temperature. At constant volume, the corresponding pressure rise with methane as the major product varies from about four to six times when the starting pressure is about 2000 atm.

Although both ethylene and propylene can decompose explosively, propylene usually requires severe conditions to decompose while ethylene is usually handled under conditions where such an event can be initiated (Britton *et al.*, 1986). For example, ethylene may react with contaminants such as oxygen and hydrogen, which can lead to a thermal runaway and finally a decomposition explosion. A thermal runaway is a result of thermal rate acceleration of an exothermic reaction system from any temperature at which the rate of heat gain exceeds the rate of heat loss from the system. It is the principal initiator of decomposition in heated equipments. Alternatively, ethylene may directly decompose by a point source of energy, by a flame, by sudden compression in presence of a diatomic gas, or a sudden excess of catalyst. Since decomposition occurs over a very short time interval (on the order of seconds), there is no practical way to control the pressure and temperature once a decomposition has initiated. To protect the reactor from bursting, the vessels are protected by bursting discs which are designed to fail at a pressure slightly above the normal working pressure. The reactor contents are safely and quickly discharged to atmosphere via blast pipes if a decomposition reaction develops and the pressure and temperature begin to rise.

Decomposition, which is commonly known as "decomp" in the polymer industry, occurs less frequently in tubular reactors because of their larger cooling surface. Autoclave reactors, however, experience such accidents sporadically. While the advanced control systems used in most polymerization reactors and the industrial product recipes guarantee the production of large volume, low cost polymers, the threat of experiencing a thermal runaway is always present for autoclave reactors. It is hard to predict decomposition because of extremely fast dynamics of such events (Zhang *et al.*, 1996).

5.3 Principal Component Analysis: Theory

Principal component analysis (PCA) is a data based multivariate statistical technique, which was developed primarily to explain the variance-covariance structure of a set of correlated variables through a few linear combinations of these variables. Usually data collected from most industrial processes contain redundancy due to multiple measurements of the same variable or due to constraints or linear relationships between variables. PCA separates this redundancy by decomposing the data into a few key independent components, which will describe the major trends in the process, thereby reducing the number of variables to be monitored.

Let $X \in \Re^{N \times n}$ represent a data matrix consisting of N measurements of n variables. The basic idea of PCA is to find n linear combinations of the original variables. By construction, these linear combinations (also known as latent or virtual variables, principal components, or scores) are uncorrelated to each other. If the original n variables are correlated, then it is possible to summarize most of the variability present in the n-variable space in terms of a lower k-dimensional subspace (k<n). If a substantial amount of the variability present in the original data set is accounted for by these k new variables (or principal components), then only these principal components can be retained for further analysis. Let us assume that we are interested in forming the following linear combinations:

$$
\begin{aligned}
t_1 &= p_{1,1}x_1 + p_{1,2}x_2 + \cdots + p_{1,n}x_n = Xp_1 \\
t_2 &= p_{2,1}x_1 + p_{2,2}x_2 + \cdots + p_{2,n}x_n = Xp_2 \\
&\vdots \\
t_n &= p_{n,1}x_1 + p_{n,2}x_2 + \cdots + p_{n,n}x_n = Xp_n
\end{aligned}
\tag{5-1}
$$

In compact form, the above system of equations can be represented as $T = XP$, where $P = [p_1 \mid p_2 \mid \cdots \mid p_n]$. Here, t_i's are known as the principal components and p_i's are known as the loading vector, which represent the weights attached to each X variable. Each loading vector is calculated such that the variance in corresponding principal component is maximized. The first principal component t_1 accounts for the maximum variance in the data, the second principal component t_2 accounts for the maximum variance that has not been accounted for by the first principal component, and so on. Some constraints need to be placed to the loading vectors to achieve this objective. These are given by

$$p_i^T p_i = 1; (\text{i} = 1,2, \ldots, \text{n}) \tag{5-2}$$

and

$$p_i^T p_k = 0; (i \neq k) \tag{5-3}$$

The first constraint is used to fix the scale of the new variable since it is possible to increase the variance of a linear combination just by scaling the weights. The second constraint ensures the orthogonality of the principal components. Given the data matrix X, the mathematical problem of calculating the weight vectors p_1 to p_n can be approached in terms of the following optimization problem:

Objective: Find a linear combination of the X variables that has the maximum variance amongst all possible linear combinations.

Objective function: $\max\{p_1^T \Sigma_{xx} p_1\}$, where $\Sigma_{xx} = X^T X$

Constraint: $p_1^T p_1 = 1$

Solution: p_1 is the first left singular vector of $\Sigma_{xx}^{1/2}$

Note that p_1 explained portion of X is represented as $Xp_1p_1^T = t_1p_1^T$, which when removed from X produces a deflated matrix E_1. Therefore, E_1 is constructed by extracting the information captured by the first PC from the original data as follows:

$$E_1 = X - t_1p_1^T \tag{5-4}$$

E_1 is treated as X to construct the second principal component $t_2 = Xp_2$. t_2 is the linear combination of the X variables which has the next highest variance subject to the conditions of being orthogonal to the first linear combination. It can be shown that p_2 is the second left singular vector of $\Sigma_{xx}^{1/2}$. The process is repeated until all the loading and score vectors are extracted. Finally, when we obtain *n* orthogonal combinations, we may write,

$$X = t_1p_1^T + t_2p_2^T + ...t_np_n^T \tag{5-5}$$

Thus, we can consider PCA as a technique that decomposes a data matrtix X into a sum of rank-one matrices. In practice, the first few scores are sufficient to capture most of the variability in the data. Therefore,

$$X = t_1p_1^T + t_2p_2^T + ...t_kp_k^T + E_k = T_kP_k^T + E_k \tag{5-6}$$

While writing the above equation, it is assumed that all the insignificant information in the data set is confined to the matrix E (which lumps the PCA dimensions k+1 through n). It is sufficient to monitor only the first *k* scores, thereby reducing the number of original variables for monitoring purpose. Therefore, one main idea behind PCA is data compression and knowledge extraction.

The loading matrix P can be calculated from a singular value decomposition (see Appendix 1) of the covariance matrix $S = X^TX/(N-1)$ as follows:

$$S = P\Lambda P^T; \quad P \in \Re^{m \times n}; \quad \Lambda \in \Re^{m \times n} \tag{5-7}$$

The loading vectors are the orthonormal column vector of the matrix P^T. Λ is the covariance matrix of the principal components. This is a diagonal matrix which contains the nonnegative real eigenvalues of decreasing magnitude ($\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_n \geq 0$) along its main diagonal with zero off-diagonal elements. PCA calculation using SVD on the

covariance matrix is equivalent to performing singular value decomposition (SVD) on the original data matrix X (Chiang *et al.*, 2000) such that

$$X = U\Lambda V^T; \quad U \in \Re^{N\times N}; \quad \Lambda \in \Re^{N\times n}; \quad V \in \Re^{n\times n} \tag{5-8}$$

The loading vectors are in this case the orthonormal eigenvectors in the columns of V^T. By construction, since loading vectors are orthogonal to each other, the scores are orthogonal or uncorrelated.

5.3.1 Fault detection using PCA

The PCA model constructed using data collected from normal plant operation can be used to perform online monitoring of the process. The "normal" PCA model of the process is used as a reference against which future plant operations are compared. Graphical tools such as the scores plot, loading plots and contribution plots are commonly used to determine the out-of-control status (fault detection) and the underlying causes for the abnormal event (fault isolation). In addition, two collective test statistics known as the Hotelling T^2 and the Q statistics are calculated from the data and are used for monitoring. A brief overview of these tools is provided in the following sections.

Scores plots: The scores plot is a description of the principal component scores for any two PCA dimensions (e.g. t_1 versus t_2). They form a two dimensional monitoring chart which indicate the relationship between the various samples. Two similar samples by virtue of their similar scores will lie close to each other in the scores plot. It is easy to conclude that all data points that are similar in nature tend to cluster together in the scores plot. The control limit contour depicting the normal operating region is an ellipse (joint confidence region). Any abnormal shift in the process variables (whether the basic correlations between variables remain intact or not) is clearly indicated in the scores plot because the projected scores move out of the operating zone. The scores plots are thus excellent tools to detect abnormal process behavior.

The scores for all n principal component can be plotted against each other. However, plotting them for the first few components is usually adequate. Once an abnormal

situation is detected by the scores plots, a loadings plot (e.g. p_1 versus p_2) may be sued for fault isolation.

SPE or the Q Statistic: If no abnormal shift occurs in the process but the correlation structure breaks down, the scores plot will not be able to detect the fault. To avoid this, a third dimension showing the squared prediction error (SPE) is calculated. In order to compute the SPE it is necessary to define the error which can also be viewed as the model plant mismatch. If this mismatch gets larger it is an indication that the PCA model no longer reflects the current status of the plant. If the process variables deviate from normal values but retain the correlation structure found during acceptable plant operation, then the SPE will not be large. It is apparent that the SPE is a useful measure to detect process upsets.

Let x_{new} denote a new multivariate observation (a vector of dimension $1 \times n$). This observation can be projected onto the hyperplane defined by the PCA loading vectors to obtain the score value $t_{new} = x_{new}P$. t_{new} is the $1 \times n$ vector of scores from the model and P is the $n \times k$ matrix of loadings determined from the normal plant data. The PCA model prediction for x_{new} is given by $\hat{x}_{new} = t_{new}P^T = x_{new}PP^T$. The $1 \times n$ dimensional error vector is given by $e_{new} = x_{new} - \hat{x}_{new}$ from which the SPE can be calculated as $e_{new}^T e_{new}$. The SPE can be considered as a scalar measure of the plant model mismatch. A small value of SPE indicates that the model is still a good representation of the plant. A large SPE value indicates otherwise. Should a large SPE value occur, the process operators should be notified, and the fault diagnostics procedure must be initiated.

The SPE, which is also known as the Q-Statistic or the Rao-statistic, employs the portion of the observation space corresponding to the (n-k) smallest singular values. The distribution of the Q-statistic has been approximated by Jackson and Mudholkar (1979):

$$Q_\alpha = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0} \qquad (5\text{-}9)$$

Where, $\theta_i = \sum_{j=k+1}^{n} \sigma_j^{2i}$ with σ_j obtained from a singular value decomposition of the data matrix X, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$, and c_α is the normal deviate corresponding to the (1- α) percentile. Given a level of significance α, the threshold for the Q-statistic can be computed using Eq. 5-9 and be used to detect faults. The Q statistic measures the random variations of the process, for example, that associated with measurement noise. A violation of the threshold would indicate that the random noise has significantly changed.

It may be added that the n principal components can be plotted individually and used as a monitoring chart in a manner similar to the univariate Shewhart chart. In this case, each of the charts represents monitoring of a group of original variables rather than a single process variable. By examining the n individual score plots and with a knowledge of the loading matrix (or a loading plot), it is possible to pinpoint the cause(s) for the abnormality. In this case, SPE must be monitored individually using a univariate procedure.

Loading plots: To help in isolating the reasons for abnormalities in process operation, it is necessary to interrogate the underlying PCA model. Some of these methods are discussed in MacGregor *et al.*, (1994). One common fault isolation technique is the use of the loadings plot which shows the relationship between the process variables in exactly the same way as the scores plot exhibits the relationship between the observations. All variables sharing the same information content (i.e. correlated variables) tend to cluster together. Such clusters of variables usually dominate different PCA dimensions.

If abnormal scores are noticed for any particular PCA dimension, the variable cluster(s) that dominate the dimension may be responsible for the unusual event. Loadings plots help in visualizing these variable clusters.

Contribution plots: The SPE values computed above can also be utilized in an effective manner for fault isolation. The fractional contribution of each process variable to the overall SPE can be computed as

$$\gamma_i = \frac{SPE_i}{SPE}; \text{ (i =1,2,...,n)} \tag{5-10}$$

where SPE_i denoted the square of the i^{th} element of the error vector e_{new}. If the fractional contribution of any variable is significant then it is very likely the cause for the abnormality. This is the basis of the contribution plot concept proposed by Miller *et al.*, (1993). It should be remembered that although an unambiguous answer regarding the source of the fault is not provided by either the loadings or the contribution plots, they definitely provide a focal point for detecting the possible causes.

T^2 Statistic: The overall measure of variability in a process is given by a quantity which is defined as:

$$T^2 = (x_i - \bar{x})S^{-1}(x_i - \bar{x})^T \tag{5-11}$$

Where $x_i \in \Re^{m\times 1}$ is a vector of n variables observed at time instant i, and S is the covariance calculated from the data. The statistic (also known as Hotelling T^2) is a quantity indicating the overall conformance of an individual observation vector to its mean or an established standard. This is a multivariate generalization of the Student t-test and gives a single answer to the question "is the process in control?" Assuming that the observations are randomly sampled from a multivariate normal distribution, and known population mean and covariance, the T^2 statistic follows a χ^2 distribution with n degrees of freedom. Then, the threshold for T^2 is given by,

$$T^2_{UCL}(\alpha) = \chi^2_\alpha(n) \tag{5-12}$$

When the actual covariance matrix is unknown and is estimated from the historical data, the upper control limit of the T^2-statistic is derived as follows:

$$T^2_{UCL}(\alpha) = \frac{n(N-1)(N+1)}{N(N-n)} F_\alpha(n, N-n) \tag{5-13}$$

Where $F_\alpha(n, N-n)$ is the 100(1-α)% critical point of the F distribution with n and (N-n) degrees of freedom (MacGregor and Kourti, 1995). Because the covariance matrix can be

ill-conditioned for correlation and collinearity between the variables, the T^2-statistic is also calculated using the first few scores and eigenvalues. Assuming data to be mean centered T^2-statistic is given by (Imtiaz *et al.*, 2006),

$$T^2 = t_r^T \Lambda_r t_r \tag{5-14}$$

where Λ_r is a diagonal matrix containing the r largest eigenvalues λ_i. T^2 statistic is therefore not affected by the inaccuracies of the smaller eigenvalues and better suited to represent the normal behavior of the process.

The T^2 and Q statistics along with their appropriate thresholds detect different types of faults, and the advantages of both statistics can be utilized by employing the two measures together. To simplify the fault detection task Raich and Cinar (1996) suggested the following combined statistic

$$CI = k\frac{SPE(x)}{Q_\alpha} + (1-k)\frac{T^2(x)}{T^2_{UCL}(\alpha)} \tag{5-15}$$

where $k \in (0,1)$ is a constant. Therefore if the statistic is less than 1 the process is considered to be normal. Excellent reviews on the theory and application of PCA can be found in Jackson (1991), and Chiang *et al.*, (2001).

5.3.2 PCA based Model Identification

In industrial data, process variables may often be related through some constraints or linear relationships such as a component mass balance or energy balance equation. Consider $X \in \Re^{N\times n}$ represent a set of N measurement on n variables. Let these variables be related through a set of m independent linear constraints such that

$$AX = 0, \tag{5-16}$$

where $A \in \Re^{m\times n}$ is a constraint matrix. To identify A is equivalent to finding the null space of the data matrix. This is also the non trivial solution of Eq. 5-16. Due to the presence of the constraints, X is a rank deficient matrix with a rank $r = n - m$, i.e., X spans a n-m dimensional subspace of R_n (denoted as V_x). Furthermore, the rows of A span a m dimensional subspace of R_n denoted as V_c which is orthogonal to V_x. Therefore if R_n can be decomposed into two orthogonal subspaces, one of which defines V_x and the other

V_c, this would solve the identification problem. Singular value decomposition of X will result in:

$$X = U\Lambda V^T; \quad where, U \in \Re^{N\times N}; \Lambda \in \Re^{N\times n}; V \in \Re^{n\times n}; \qquad (5\text{-}17)$$

The diagonal elements of the matrix Λ will have entries $\lambda_1^2, \lambda_2^2, ..., \lambda_r^2 > 0 = \lambda_{r+1}^2, \lambda_{r+2}^2, ..., \lambda_n^2$. The transpose of the right singular vectors associated with the zero singular values define a basis for V_c and is a solution for the constraints matrix A.

PCA can be applied to solve the above problem by finding the orthonormal eigenvectors of the covariance matrix S. It is easy to observe that S has a rank of ***n-m***. Thus it will have ***n-m*** non-zero eigenvalues while the rest of the eigenvalues are zero. The model is obtained from the eigenvectors associated with the zero eigenvalues.

It is well known that an exact model can be identified using the above theory only in the ideal case, when the measurements are not corrupted with noise. Random errors in the measurements will convert S into a full rank matrix and also convert the zero eigenvalues to non zero but small eigenvalues. However, if the noise level is low, and signal to noise ratio (SNR) is high, then we can still use this method by selecting the eigenvector corresponding to the small non zero eigenvalues. The following simple simulation example illustrates the identification method.

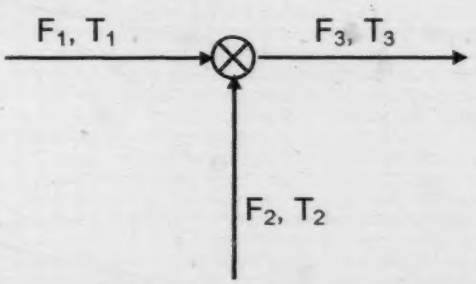


Figure 5-2: Model Identification for a mixing node

Consider a simple example where two streams of liquid water having flow rates F_1, F_2, and temperatures T_1 and T_2 respectively are mixed. The product stream has a flow rate F_3 and temperature T_3. An energy balance for the mixing node may be written as

$$F_1 C_p (T_1-T_{ref}) + F_2 C_p (T_2-T_{ref}) = F_3 C_p (T_3-T_{ref}) \quad (5\text{-}18)$$

Assuming $T_{ref} = 0$, and constant heat capacity C_p for water at moderate temperature and pressure, Eq. 5-18 reduces to:

$$F_1T_1 + F_2T_2 = F_3T_3 \quad (5\text{-}19)$$

Three sets of data with different signal to noise ratio were generated for the above example. Each data set had 1000 realizations of F_1, F_2, F_3, T_2, T_2, and T_3 where the following mass and energy balance equations hold:

$$F_3 = F_1 + F_2$$

$$T_3 = (F_1T_1 + F_2T_2)/F_3$$

In order to simulate the true values of variables at each time instant, a set of independent flow (F_1 and F_2) and temperature (T_1 and T_2) variables were chosen. The true values of independent variables were simulated by adding normally distributed random fluctuations to their base values. The true values of the dependent variables (F3 and T3) were calculated such that they satisfy the mass and energy balance equations. The mean values of variables and the standard deviations of the fluctuations are given in Table 5-1. In the simulation procedure, the measured values of variables were simulated by adding normally distributed random noise to their true values. The standard deviations of measurement errors are also given in Table 5-1. A measure of signal to noise ratio (SNR) is given by the ratio of standard deviation of the signal fluctuations to the standard deviation of noise.

Next, PCA was applied to: $X = [F_1T_1F_2T_2 \quad F_3T_3]$. The results of PCA analysis are summarized in Table 5-1.

It should be noted that PCA is a scale dependent technique and often requires proper scaling in order to obtain reasonable results. Particularly when the variables have different units and vary over a wide range of magnitude, it is common practice to remove the means from the variables and scale with the standard deviation of the data. It is shown by Narasimhan and Shah (2007) that in the noise-free case PCA can result in an exact basis for V_c, even if the data is auto-scaled. However, if we apply scaling to measurements corrupted with noise, then no simple relationship can be obtained between the eigenvectors of the scaled covariance matrix and those of the unscaled raw data, and model identification is not at all straightforward. Therefore, when applying this theory to real industrial data, one has to be careful about the level of excitement present in the data, the source of the variables in the process and their physical significance. If the signal to noise ratio is high and the variables have comparable units and magnitudes, it is better to do the analysis with unscaled data.

Table 5-1: PCA based model identification using noisy data

SNR	Variable	True values		Std of noise	Eigen-values	Last eigen-vector	Comment
		Mean	Std of fluctuations				
50 (high)	F1	10	1	0.1	2490000	0.57	Exact model found
	F2	20	1	0.1	715	0.57	
	F3	-	-	0.1	0.76	-0.57	
	T1	20	1	0.1			
	T2	50	1	0.1			
	T3		0	0.1			
10 (medium)	F1	10	1.00	0.5	2490000	-0.59	Close to exact model
	F2	20	0.99	0.5	750	-0.57	
	F3	-	1.00	0.5	19.5	0.57	
	T1	20	0.99	0.5			
	T2	50	0.96	0.5			
	T3		0.98	0.5			
2 (low)	F1	10	1	0.5	2490000	-0.83	Exact model not found
	F2	20	1	0.5	1150	-0.35	
	F3	-	-	0.5	386	0.43	
	T1	20	1	0.5			
	T2	50	1	0.5			
	T3	-	-	0.5			

5.4 Process description

An industrial LDPE/EVA autoclave reactor at AT Plastics (Edmonton, Canada) was studied for this project. A brief description of the process was given in section 3.3 of this thesis. The plant uses ICI high-pressure technology for the polymerization of ethylene to LDPE and copolymerization of ethylene and vinyl acetate to EVA copolymers by free radical polymerization. Figure 5-3A shows a photograph of the reactor before it was installed. Figure 5-3B shows a cross section of the same reactor. A schematic of feed and initiator flow arrangement is shown in Figure 5-3C.

The reactor vessel is cylindrical in shape with a nominal capacity of 750L and a length to diameter ratio of 10:1. It is equipped with an axial stirrer and thermocouples along the reactor length. The stirrer ensures good mixing and uniform heat distribution in the

reaction mixture. The reaction space is divided into four separate reaction zones by means of three baffles attached to the stirrer. Ethylene gas is supplied to the reactor from the secondary compressor through four separate gas steams. Free radical initiators are injected continuously into each zone of the reactor to control the exothermic polymerization reaction at selected temperature levels. The reactor operates at a pressure in the range of 1000 to 3000 atm and achieves about 20% conversion in a single pass. After separation from the unreacted gases in the high pressure separator and extrusion hopper, molten polymer is sent to the extrusion, drying and packaging units. Ethylene gas separated from the high pressure separator is recycled to the suction of the secondary compressor. Any entrapped ethylene separated from the extrusion hopper is compressed in a booster compressor, and finally mixed with fresh ethylene at the suction of the primary compressor. The same unit also produces ethylene vinyl acetate copolymers (EVA) when vinyl acetate is added to the system.

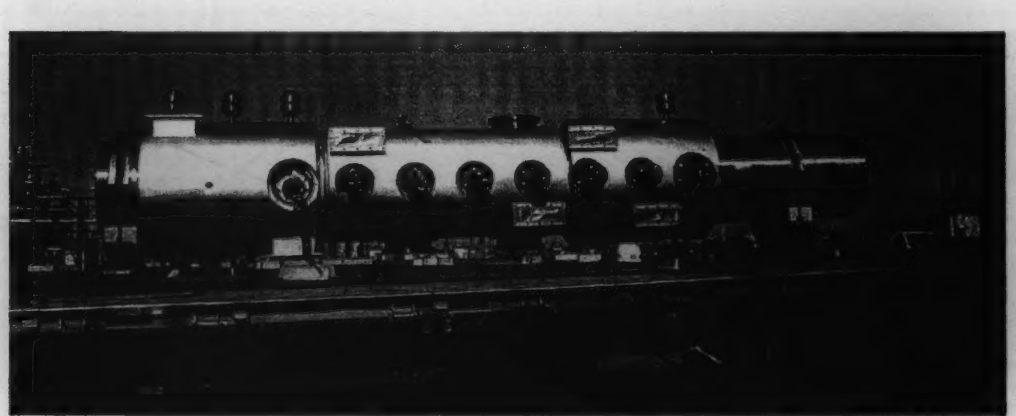


Figure 5-3A: Autoclave reactor at AT Plastics (note the scale by comparing the van in front of the reactor)

To ensure safe operation of the reactor vessel, some critical process variables are constantly being monitored by the operators. These include the reactor zone temperatures, pressure, and catalyst pump loading. In addition, the feed temperature, feed gas pressure, feed composition, polymer melt temperature at the reactor exit etc. are also recorded. Composition of the feed gas is analyzed using a gas chromatograph before it enters the reactor. These variables along with process variables from all other units at the plant are stored in a data historian. The plant uses a Honeywell TDC 3000 DCS and an Aspen Tech's Infoplus21 historian to collect and store all the process data. Almost all of the critical variables such as pressure, temperature and fresh feed flow rates etc. are sampled at every second and stored in the historian without any compression. The historian is large enough to store three years of measured process data without compression.

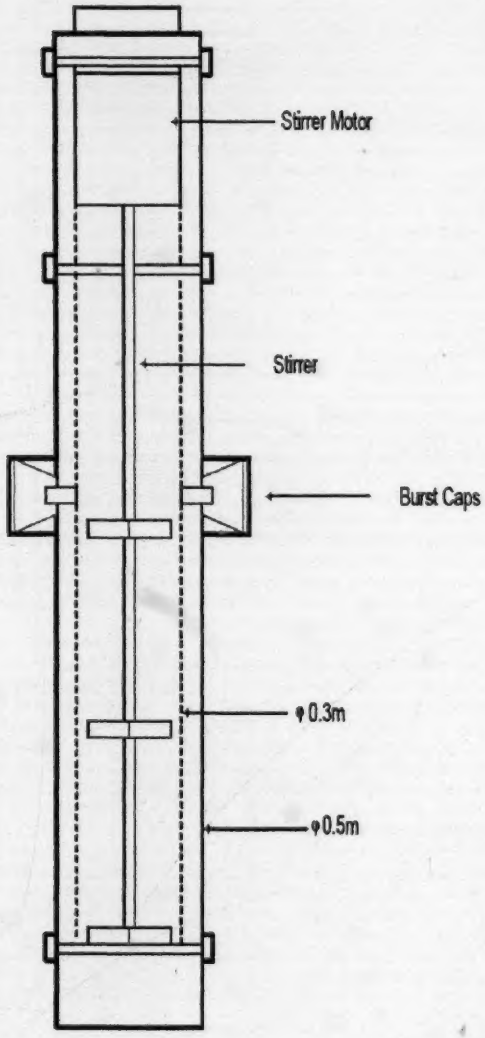


Figure 5-3B: Cross Section of the Reactor

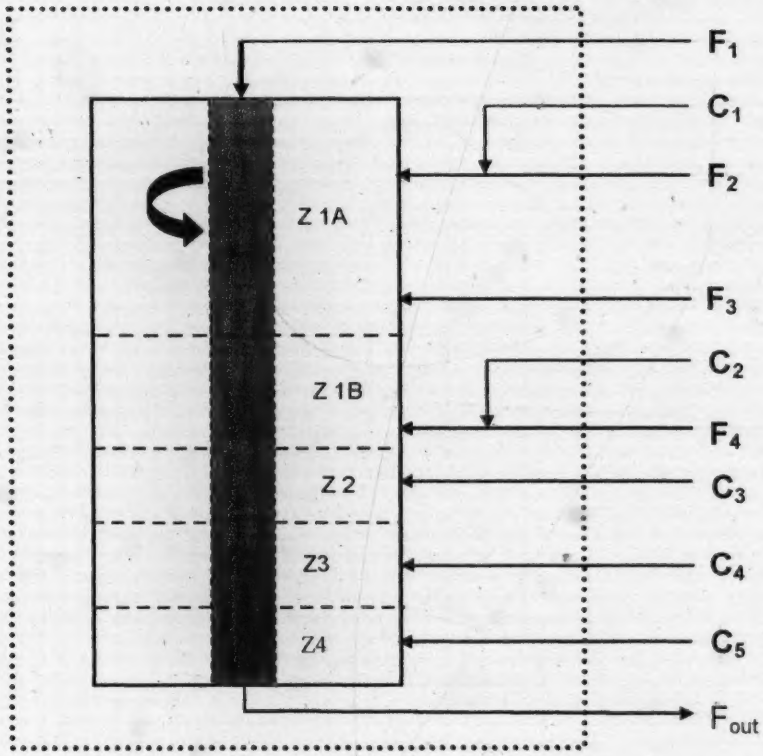


Figure 5-3C: Schematic of the Autoclave reactor configuration

5.5 Formulation of Overall Heat Balance around the Reactor

Polymerization of ethylene is an exothermic reaction, which generates heat. In order to maintain a sufficiently high reaction rate, the feed gas needs to be heated to the reaction temperature. During start up, the reactor is heated externally using hot air, but as the reaction proceeds, heat released from polymerization is absorbed by the cold incoming feed gas. In addition to this, viscous dissipation from the reaction mixture due to the rotation of the stirrer generates heat. The bottom zones of the reactor wall are surrounded by a hot oil jacket. Temperature of the hot oil is maintained approximately at the same level as the reactor temperature, hence is may be assumed that heat loss through the bottom zones are negligible. However, a small amount of heat is lost from the top zone of the reactor, which is not surrounded by the oil jacket. Therefore, heat removal from the reactor is achieved mainly by sensible heat transfer to the feed gas and the products, and to a smaller extent, by heat loss through the wall. During steady state, the reactor operation can be considered as auto thermal, where the main source of heat generation inside the reactor is the polymerization reaction, and the principal means of heat removal is by the sensible heat absorbed by the cold feed and the products.

To formulate a steady state overall energy balance for the reactor, consider the energy balance equation for any open system at steady state as a starting point:

$$\Delta\dot{H} + \Delta E_K + \Delta E_P = Q - W_s \qquad (5\text{-}20)$$

here,

$\Delta\dot{H}$ = rate of change of enthalpy of the reacting mixture

ΔE_K = rate of change of kinetic energy; since the linear velocities of the inlet and exit streams are similar (typical inlet and exit stream velocities are 11.14 and 12.03 m/s at inlet and exit respectively) ΔE_K is neglected

ΔE_P = rate of change of potential energy; since there is no significant change of elevation, we can neglect ΔE_P

Q = rate of heat transfer to the system from the surroundings

W_s = shaft work done by the system on the surroundings, since no shaft work is considered for the reactor W_s is neglected. (The compressor is pressurizing the gas into the reactor but we consider only the reactor as the system).

The following steps are required as part of the standard procedure of calculating $\Delta\dot{H}$ for any reactive system:

1. Choose reference conditions: If heat of polymerization reaction is known at temperature T_r and pressure P_r, then choose all compounds at the inlet and the outlet of the reactor at T_r and P_r as the reference condition
2. Calculate extent of reaction ξ from mass balance
3. Calculate specific enthalpy of each component $\hat{H}_i$ in the inlet and the outlet stream relative to the reference condition. If the effect of pressure on C_p is ignored, then $\hat{H}_i$ is calculated as $\hat{H}_i = \int_{T_r}^{T} C_p dT$
4. Calculate $\Delta\dot{H}$ using: $\Delta\dot{H} = \xi\Delta\hat{H}_r^0 + \sum \dot{n}_{out}\hat{H}_{out} - \sum \dot{n}_{in}\hat{H}_{in}$. Here, $\Delta\hat{H}_r^0$ is the heat of reaction at reference condition, $\hat{H}_{in}$ and $\hat{H}_{out}$ are specific enthalpies of each component in the inlet and outlet streams relative to the reference conditions, and n is molar or mass flow rates of components

For most of the polymer grades produced at the plant, the major reactants are ethylene, VA, and propylene. The reactor outlet stream mainly contains polymer (LDPE or EVA), and unreacted ethylene and VA. In addition, liquid solutions of free radical initiator are continuously added into the reactor. The product from the reactor contains the solvent, and small amount of inert gases such as methane, ethane, and CO_2 (produced from decomposition of peroxide initiator). However, it can be assumed that the initiators are completely consumed in the reaction and the masses of solvents and initiators are usually small compared to the reactants and products. Hence they are ignored in all mass and energy balances. Considering the major components $\Delta\dot{H}$ is calculated as follows:

$$\Delta\dot{H} = \xi\Delta\hat{H}_r^0 + \sum \dot{n}_{out}\hat{H}_{out} - \sum \dot{n}_{in}\hat{H}_{in}$$

$$= \xi\Delta\hat{H}_r^0 + (\dot{m}_{ethylenes}\hat{H}_{ethylene} + \dot{m}_{VA}\hat{H}_{VA} + \dot{m}_{EVA}\hat{H}_{EVA})_{exit}$$

$$- (\dot{m}_{ethylene}\hat{H}_{ethylene} + \dot{m}_{VA}\hat{H}_{VA} + \dot{m}_{propylene}\hat{H}_{propylene})_{inlet}$$

$$= \xi\Delta\hat{H}_r^0 + (\dot{m}_{ethylenes}\int_{Tr}^{T} Cp_{ethylene}dT + \dot{m}_{VA}\int_{Tr}^{T} Cp_{VA}dT + \dot{m}_{EVA}\int_{Tr}^{T} Cp_{EVA}dT)_{exit}$$

$$- (\dot{m}_{ethylenes}\int_{Tr}^{T} Cp_{ethylene}dT + \dot{m}_{VA}\int_{Tr}^{T} Cp_{VA}dT + \dot{m}_{propylene}\int_{Tr}^{T} Cp_{propylene}dT)_{inlet}$$

(5-21)

Defining Q_{wall} as the rate of heat loss through reactor wall, and $Q_{stirrer}$ as the rate of heat generation by the stirrer, and using Eq-18, the final form of the balance equation is as follows:

$$\xi\Delta\hat{H}_r^0 + (\dot{m}_{ethylenes}\int_{Tr}^{T} Cp_{ethylene}dT + \dot{m}_{VA}\int_{Tr}^{T} Cp_{VA}dT + \dot{m}_{EVA}\int_{Tr}^{T} Cp_{EVA}dT)_{exit}$$

$$- (\dot{m}_{ethylenes}\int_{Tr}^{T} Cp_{ethylene}dT + \dot{m}_{VA}\int_{Tr}^{T} Cp_{VA}dT + \dot{m}_{propylene}\int_{Tr}^{T} Cp_{propylene}dT)_{inlet} = Q_{stirrer} - Q_{wall}$$

(5-22)

5.6 Order of Magnitude Analysis

In order to determine the relative importance of the terms in Eq. 5-22, we performed an order of magnitude analysis by developing a simplified steady state energy balance for the production of LDPE. In this case, the feed stream contains only ethylene and the product stream contains LDPE and unreacted ethylene. For simplicity, it was assumed that the heat capacities of ethylene and polyethylene are the same and constant over the temperature range inside the reactor. $\Delta\dot{H}$ is estimated as:

$$\Delta\dot{H} = \xi\Delta\hat{H}_r^0 + \sum \dot{n}_{out}\hat{H}_{out} - \sum \dot{n}_{in}\hat{H}_{in}$$

$$= \xi\Delta\hat{H}_r^0 + \dot{m}_{feedgas}\hat{H}_{product} - \dot{m}_{feedgas}\hat{H}_{feedgas}$$

$$= \xi\Delta\hat{H}_r^0 + \dot{m}_{feedgas}C_{p,ethylene}(T_{reactor_exit} - T_{ref}) - \dot{m}_{feedgas}C_{p,ethylene}(T_{feedgas} - T_{ref})$$

$$\therefore \Delta\dot{H} = \xi\Delta\hat{H}_r^0 + \dot{m}_{feedgas}C_{p,ethylene}(T_{reactor_exit} - T_{feedgas}) \qquad (5\text{-}23)$$

Extent of reaction ξ is defined as the total mass of ethylene that takes part in the reaction. If X represents single pass fractional conversion, then $\xi = \dot{m}_{feedgas}X$. Using this relationship, we obtain,

$$\dot{m}_{feedgas}X\Delta\hat{H}_r^0 + \dot{m}_{feedgas}C_{p,ethylene}(T_{reactor_exit} - T_{feedgas}) + Q_{stirrer} - Q_{wall} = 0 \qquad (5\text{-}24)$$

In order to get an estimate of each term in Eq. 5-24, we used some average representative values for feed gas flow rate, conversion, power input to the stirrer, and temperature difference across the reactor. Using $\Delta\hat{H}_r^0$ = 807 Cal/g ethylene polymerized, $C_{p,ethylene}$ = 0.6Cal/g ^{0}C, and average thermal conductivity of the reactor wall, K_{wall} = 100 Cal/m^2 s ^{0}C, we obtained:

$\dot{m}_{feedgas}X\Delta\hat{H}_r^0 = 2.83\text{x}10^{10}$ J/hr

$\dot{m}_{feedgas}C_{p,ethylene}(T_{reactor_exit} - T_{feedgas}) = 2.17\text{x}10^{10}$ J/hr

Assuming that the power input to the stirrer roughly represents the heat generated by its rotation, $Q_{stirrer}$ = 4.5x10^8 J/hr. On an average summer day when the ambient temperature is 20 ^{0}C, Q_{wall} = 1.8 x 10^8 J/hr. On an average winter day when the ambient temperature is - 20 ^{0}C, Q_{wall} = 2.3 x 10^8 J/hr.

It can be concluded from the above analysis that heat generated by the stirrer rotation and heat loss through the reactor wall are of much smaller magnitude as compared to the heat of reaction and the sensible heat transfer terms. Also, Q_{wall} and $Q_{stirrer}$ are of similar magnitude and tend to cancel each other. Therefore, these terms can be safely ignored from the balance equation. In order to calculate the remaining terms in Eq. 5-22 on a continuous basis, the following information was required:

1. Molar (or mass) flow rates of all reactants and products
2. Heat capacity of each compound (C_2H_4, PE, EVA, VA, propylene) and their dependency with temperature and pressure
3. Heat of reaction of polymerization for different grades of EVA

The following sections briefly explain how these quantities were estimated.

Total flow rate: Mass flow rate of fresh ethylene, VA and propylene is measured at the feed point of each of these streams, but the total mass flow into the reactor is not measured in the plant. A first principles model to estimate the mass flow rate through the secondary compressor was provided by the plant personnel and was used successfully in a previous work by Kumar *et al.* (2003). The same model was used in this work to calculate the mass flow rate.

Single pass conversion: Although it was known from ICI process documents that average single pass conversion varies from 16-20%, no online measurement of this quantity was available at the plant. Two methods were employed to estimate conversion value online. The mass of total polymer produced in every batch is measured at the storage silo located further downstream from the reactor. A measure of ethylene conversion in the reactor was calculated from the total mass of polymer and total flow rate of ethylene into the reactor. However, this method will result in significant lag due to the distance between the reactor and the storage silo. An alternative way was to estimate this value from mass flow rate measurements of fresh ethylene feed and reactor feed. This method, which is explained in Figure 5-4, is based on steady state flow. In this simplified block diagram, all flow rates represent mass of ethylene. Assuming small amount of ethylene loss in the separator, conversion can be calculated from $\alpha F_1/F_3$, where α is a loss factor. Here F_1 represents fresh ethylene flow rate, which is measured in the plant. F_3 is the mass of ethylene into the reactor, which was estimated using the total mass flow and the feedgas compositions. This method gave reasonable results with negligible lag and was used for this analysis.

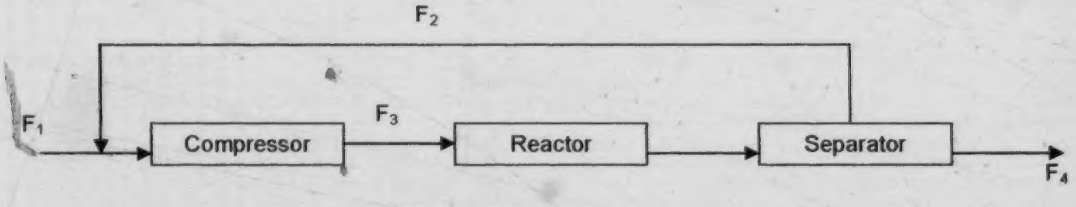


Figure 5-4: Block diagram to estimate conversion

Thermodynamic data: One difficulty encountered in this work was finding the appropriate heat capacity value of the components at supercritical conditions. This is common for many industrial processes, where accurate heat capacity data may not be readily available in standard heat capacity tables and charts. To overcome this difficulty, we used an average heat capacity of each compound, which is closest to the true value at the reactor condition. Some excellent sources of thermodynamic data for hydrocarbons used include Gaur & Wunderlich, 1981, and Nehzat, 1980. In addition, industrial databases often document thermodynamic data related to their particular process. For this work, Aspen Tech's thermodynamic database was used to obtain C_p values for the gas mixture at reactor exit conditions. However, such data base is of proprietary nature and is not available for general use. Where no reliable source was found, an average heat capacity for supercritical fluid was assumed to be the same as the corresponding ideal gas heat capacity (Perry's Chemical Engineering Handbook). Heat capacity of propylene at supercritical conditions was approximated using this method. Although this does not represent the true value, it gives a much closer estimate as compared to the corresponding liquid heat capacity. The variation of C_p with pressure and temperature was assumed unknown. There is also some uncertainty associated with heat of reaction of EVA products as this will depend on the particular grade of polymer. We used an average value of 807 Cal/g of ethylene polymerized, which is quite accurate for LDPE. To accommodate for the uncertainties, the heat balance equation is then modified, where each term in Eq. 5-22 is decomposed into a known and an unknown (or a poorly known) part:

$$\alpha_{reaction}\left(\xi\Delta\hat{H}_r^0\right)+$$
$$\left\{\left(\dot{m}_{ethylenes}\overline{C}p_{ethylene}\Delta T\right)\alpha_{ethylene}+\left(\dot{m}_{VA}\overline{C}p_{VA}\Delta T\right)\alpha_{VA}+\left(\dot{m}_{EVA}\overline{C}p_{EVA}\Delta T\right)\alpha_{EVA}\right\}_{exit}$$
$$-\left\{\left(\dot{m}_{ethylenes}\overline{C}p_{ethylene}\Delta T\right)\alpha_{ethylene}+\left(\dot{m}_{VA}\overline{C}p_{VA}\Delta T\right)\alpha_{VA}+\left(\dot{m}_{propylene}\overline{C}p_{propylene}\Delta T\right)\alpha_{propylene}\right\}_{inlet}=0 \quad (5\text{-}25)$$

By denoting the known term as H_i and the unknown (or poorly known) terms as α_i the above equation can be simplified as:

$$\alpha_1 H_1+\alpha_2 H_2+\ldots+\alpha_n H_n=0 \quad (5\text{-}26)$$

Now the problem of formulating the energy balance is reduced to solving Eq. 5-26 for the unknown α's. PCA was used as a model identification tool to find a solution

5.7 PCA Model Development

A number of data sets which included fresh feed flow rates, feed gas composition, reactor feed temperature, and temperature of each reaction zone were collected from the plant site. As explained, the total feed flow rate and the fractional conversion values were calculated and appended to the data matrix. Component mass flow rates at reactor inlet and exit were estimated from mass balance equations. Before proceeding to model building, as a standard data pre-processing technique, the data was checked for compression, quantization, and the need for filtering. Industrial data is often compressed when stored in the historian, primarily to save storage space. A measure of compression in sampled data is given by what is defined as a 'compression factor'. Thronhill *et al.*, (2004) suggested that a compression factor higher than three will destroy the variance-covariance structure of the data and should not be used for any data based analysis. However, as mentioned earlier, all recent data in the plant is stored without any compression and could be used for further analysis. The variables which were used in this analysis were mostly collected from the reactor or areas close to the reactor. Some of these variables are usually well controlled within safe limits during normal operation, hence the signal to noise ratio was sufficiently high. The feed flow rate and conversion were among the estimated variables and had higher level of noise as measurements from many different sections of the plant were required for their estimation. These variables were filtered using a low pass EWMA filter.

Normal data from three different batches of EVA production were combined to form the training data set for PCA model. PLS toolbox in MATLAB was used to perform the basic PCA calculations. All known terms in Eq. 5-25 (shown within first bracket) were calculated. An average heat capacity value for the product was used for the reactor exit stream. Therefore, in Eq. 5-25, the three terms for components in the reactor exit were condensed into one; hence the X matrix for PCA had five columns. As explained earlier, selection of appropriate number of PC is often not a straightforward task when dealing with industrial data. In this study we used the relative magnitude of eigenvalues as a criterion to choose how many PC's to retain in the model. This is equivalent to using the variance explained by each PC as the selection criteria. It was observed that the second PC was small enough to use as representative of the heat balance error. The signs of the loading vector corresponding to the 2^{nd} PC were in close agreement with Eq. 5-25. More details on selection of the appropriate PC will be given in the next section. Next, a 95% confidence interval was calculated for this score vector using the normal data which gave an upper and a lower bound for the selected PC. These limits were considered as the safe operation limit for the reactor. If the score which is an estimate of closure error exceeds the upper bound, then it was assumed that excess heat is generated inside the reactor which was not being removed by the feed streams. Figure 5-5 shows the selected score variable corresponding to PC2 that represents the heat balance error calculated from the normal data along with the estimated safety limits.

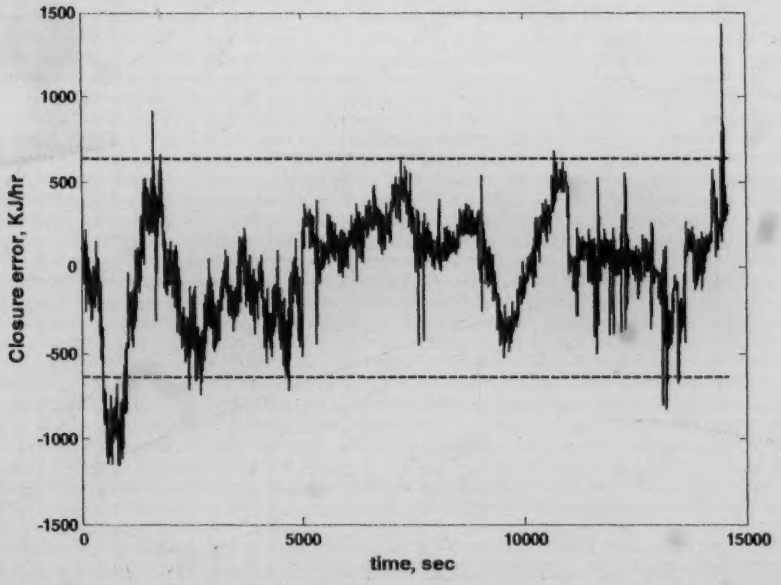


Figure 5-5: Energy balance closure error as estimated from the normal PCA model

5.8 Choice of Appropriate PC

In this section we present some results obtained from a detailed analysis on a criterion for selection of the appropriate PC. The analysis was performed with data collected from an EVA grade which did not contain any propylene. Therefore, the data matrix for PCA model development had four columns instead of five. Normal data from three batches of production of the same grade were combined to form the training data for this analysis. All standard data preprocessing steps were followed as described before. We emphasize the following points for clarity:

- No scaling was used
- An average heat capacity value was used to calculate component enthalpies
- Standard heat of ethylene homo-polymerization was used to approximate heat of EVA copolymerization
- Variation of heat capacity with temperature and pressure and variation of heat of reaction with different comonomer content (weight % of vinyl acetate) was assumed unknown

The data matrix was configured as follows: $X = [H_{reaction} \quad H_{mix} \quad H_{VA} \quad H_{ethylene}]$. Table 5-2 gives some details on how each of these terms was calculated. Average magnitudes of these terms during normal operation are also shown.

Table 5-2: Calculation of various terms in X matrix

Term	Calculation formula	Average magnitude (J/hr)
$H_{reaction}$	$F.(conversion).(\Delta H_R)$	24500
H_{mix}	$F.Cp_{mix}.T_{mix}$	21500
H_{VA}	$F.(X_{VA}).(Cp_{,VA}).T_{VA}$	100
$H_{ethylene}$	$F.(X_{ethylene}).(Cp_{,ethylene}).T_{ethylene}$	550

Here,

$H_{reaction}$ = heat released by polymerization reaction

H_{mix} = enthalpy of the product stream relative to reference condition

H_{VA} = enthalpy of vinyl acetate feed relative to reference condition

H_{ethy} = enthalpy of ethylene feed relative to reference condition

F = total mass flow through the reactor

X_i = mass fraction of different components in total flow

Plot of each column in X matrix is shown in Figure 5-6. Different segments in Figure 5-6 represent different batches of data. The data contains sufficient amount of variability, which makes them suitable to use for model identification. Closer observation of this figure shows strong correlation among the mixture, VA and ethylene enthalpies. For better comparison, the data in X was scaled by dividing each column with its maximum value such that the range of each column is between 0 and 1. Figure 5-7 shows a plot of the scaled variables in X. The correlations among H_{mix}, H_{VA}, and H_{ethy} are quite apparent in Figure 5-7.

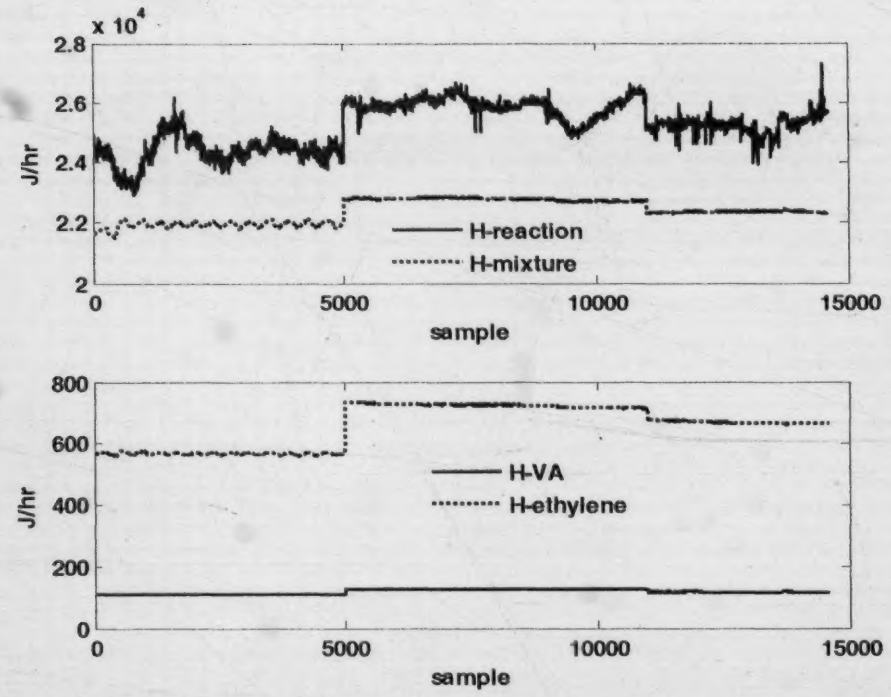


Figure 5-6: Plot of X during normal operation (used to build the PCA model)

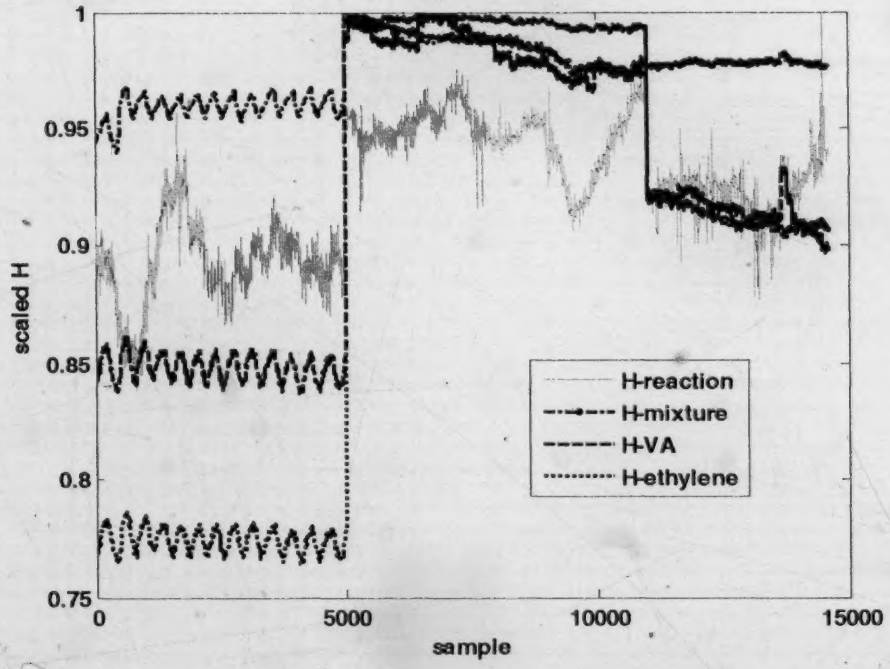


Figure 5-7: Plot of scaled X during normal operation

If we assume that all the thermodynamic parameters used in the above calculations are exact, then a theoretical value of the heat balance closure error may be obtained as follows:

$$\text{Closure Error} = H_{reaction} - [H_{mix} - H_{VA} - H_{ethy}] \qquad (5\text{-}27)$$

Note that even during normal operation, the error will be non-zero to account for heat loss terms which were neglected in developing this relationship. One data set was collected where decomposition took place. All four variables in X matrix were calculated for this new data set. The error was calculated according to Eq. 5-27 and is plotted in Figure 5-8. This figure will be used later for validating the selected PCA model.

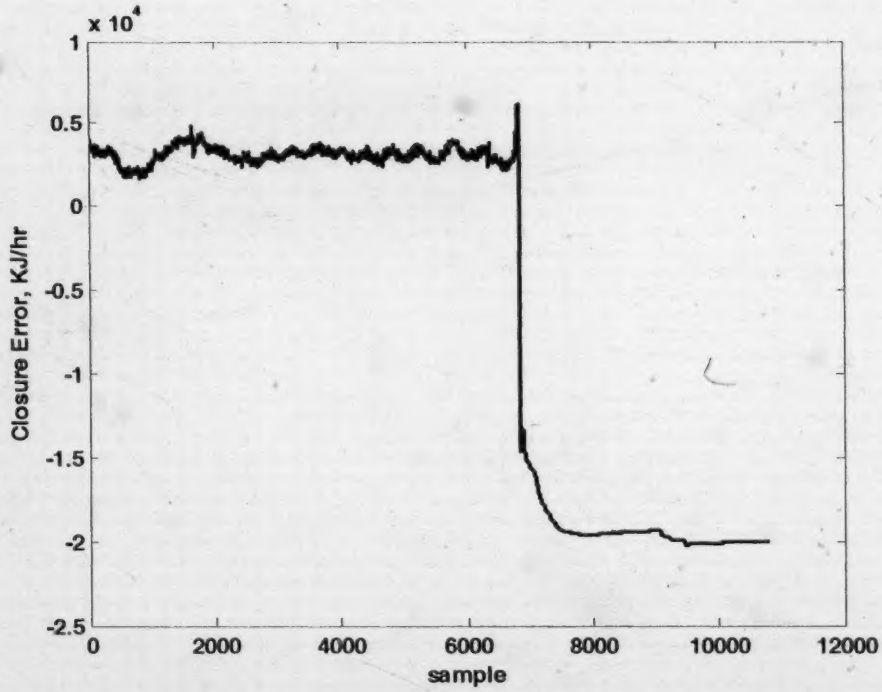


Figure 5-8: Theoretical value of closure error calculated from mechanistic model (Eq. 5-27)

Next, PCA was applied to the data matrix X. Eigenvalues of the covariance matrix of X and individual and cumulative variances explained by each principal component are shown in Table 5-3. Table 5-4 shows the loading vectors (eigenvectors of the covariance matrix) corresponding to each PC. Figures 5-9(a) to 5-9(d) show the scores calculated for the normal PCA model. It is observed that three eigenvalues corresponding to the last three PC's are quite small compared to the first eigenvalue (Figure 5-10). It is expected that only one relationship, which is the desired energy balance equation for the reactor exists among the H variables constituting the X matrix. Then theoretically we should expect to have only one "small" eigenvalue from eigen decomposition of the covariance matrix of X. However, it should be noted that in calculating the last three columns of X, the following two mass balance equations were utilized:

$$\text{Total mass balance: } F_1 + F_2 = F \tag{5-28}$$

$$\text{Component balance: } F_1 = x_1 F \tag{5-29}$$

Where F_1 and F_2 are mass flowrates of ethylene and vinyl acetate in the feed stream, F is the flowrate of the reactor product stream, and x_1 represents ethylene mass fraction in the total flow. Inclusion of these mass balance relationships in the data matrix will result in correlated component enthalpy terms. Although the purpose of this analysis is to approximate the overall heat balance relationship among all the enthalpy terms, rather than finding any explicit relationship among the component enthalpy terms, this will result in three smaller eigenvalues in the PCA analysis (as was observed in Table 5-3). The problem is now to judiciously choose the correct principal component.

Table 5-3: Results from the normal PCA Model

Principal component	Eigenvalue of covariance(X)	% Variance captured by this PC	Cumulative % variance captured
1	1.14e+009	99.99	99.99
2	9.12e+004	0.01	100.00
3	2.38e+003	0.00	100.00
4	1.43e+000	0.00	100.00

Table 5-4: Loading vectors from the normal PCA Model

p_1	p_2	p_3	p_4
-0.7476	0.6608	0.0667	0.0001
-0.6639	-0.7464	-0.0454	0.0023
-0.0035	0.0058	-0.0946	-0.9955
-0.0196	0.0781	-0.9922	0.0948

After careful investigation of the eigenvectors, the 2nd PC was selected as representative of the heat balance error for the following reasons:

1. Percentage variance captured by the 2nd PC is negligible (relative magnitude of eigenvalue corresponding to this PC is small)

2. The signs of the 2nd loading vector agree with what is expected from the theoretical model (see Eq. 5-27).
3. Relative magnitudes of the weightings on the 2nd loading vector also seem reasonable. Heat of polymerization reaction and the enthalpy of the product stream are the major contributors to the overall reactor heat balance. The enthalpy of the feed streams (ethylene and VA) should have less effect in the heat balance equation because feed temperature is lower compared to the product temperature. Enthalpy of ethylene feed stream should be higher than enthalpy of VA stream since ethylene flow rate is almost 3 times higher that VA flow rate. All of these theoretical observations are supported by the weightings of the 2nd loading vector.

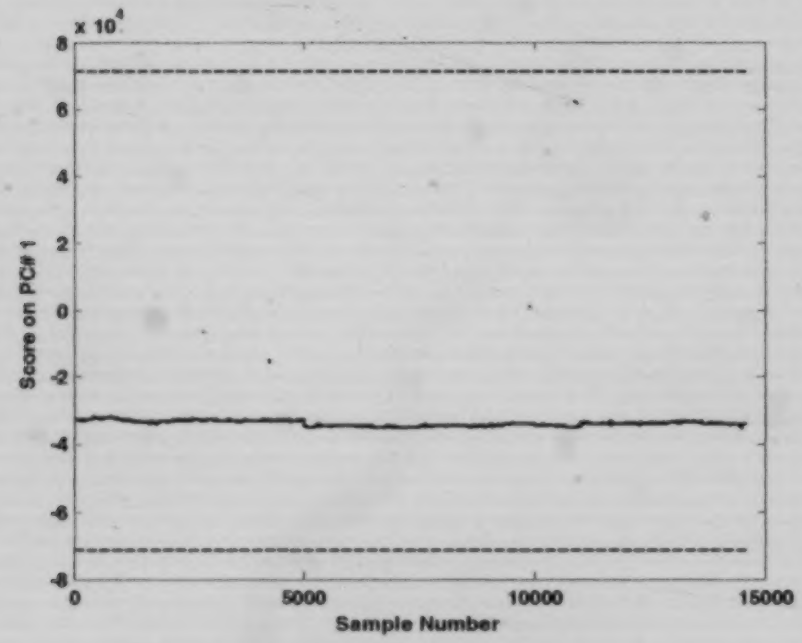


Figure 5-9(a): 1st PC from normal PCA model

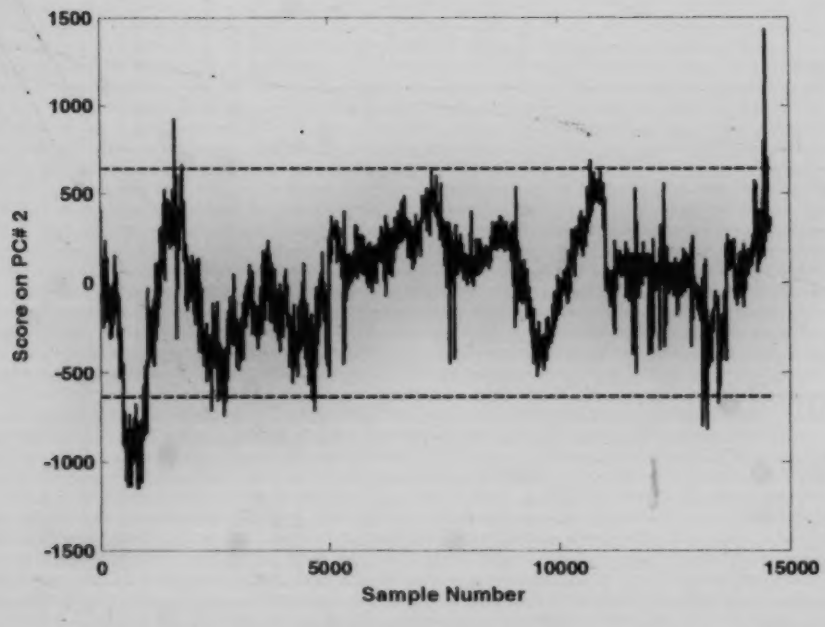


Figure 5-9(b): 2nd PC from normal PCA model; This score was used for model validation

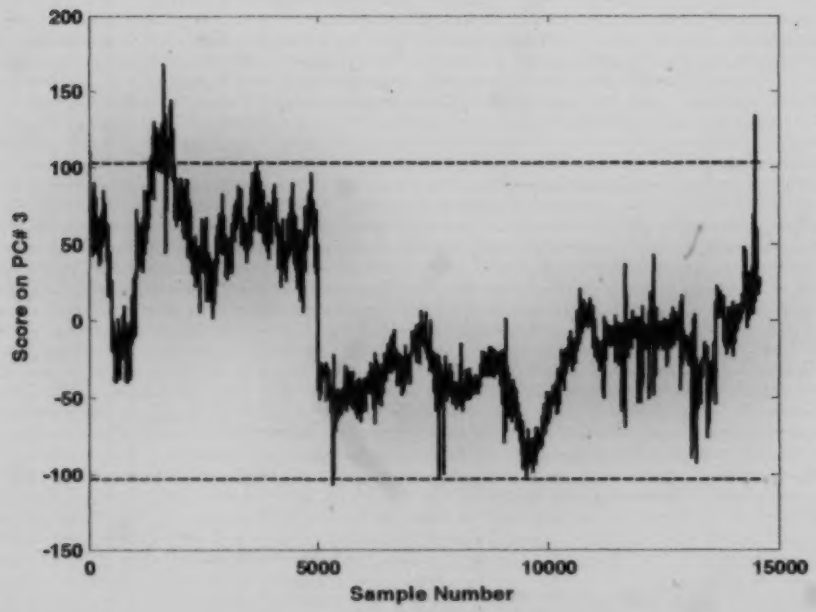


Figure 5-9(c): 3rd PC from normal PCA model

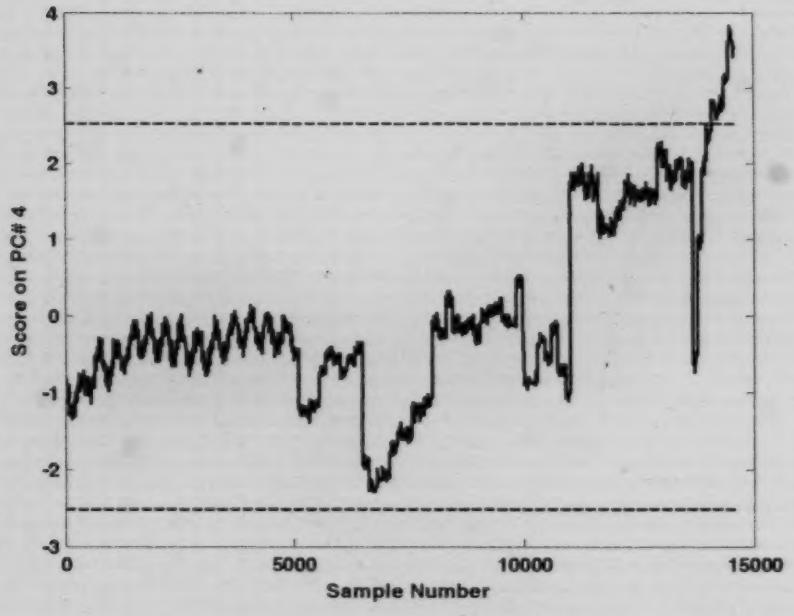


Figure 5-9(d): 4th PC from normal PCA model

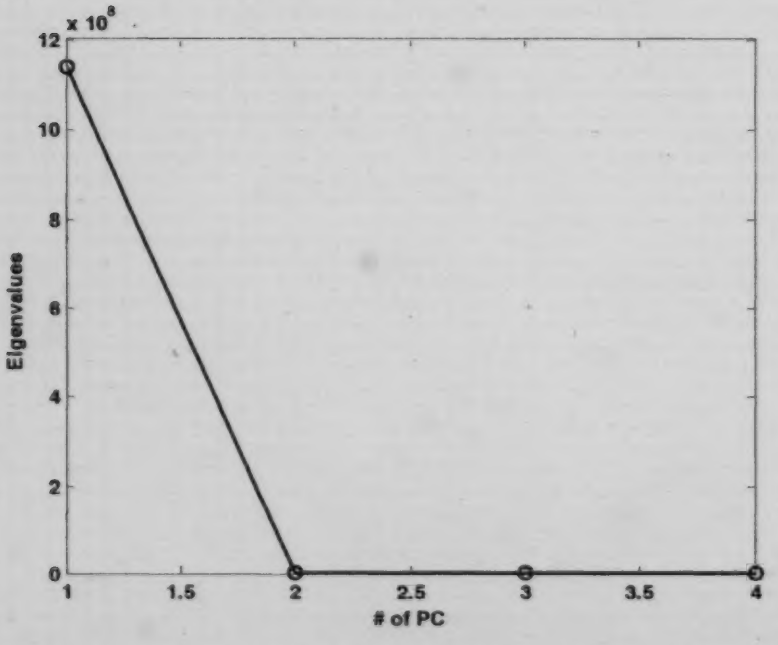


Figure 5-10: Plot of Eigenvalues of the Covariance matrix of X (Scree Plot)

Next, a new data set containing decomposition was used to validate the PCA model. The new validation scores calculated from the PCA model are shown in Figures 5-11(a) through 5-11(d)

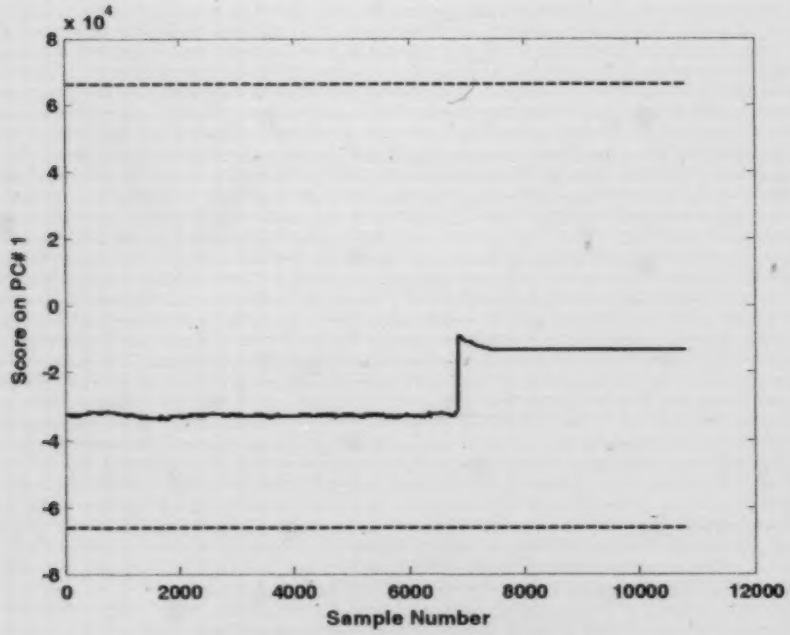


Figure 5-11(a): 1st PC from validation data

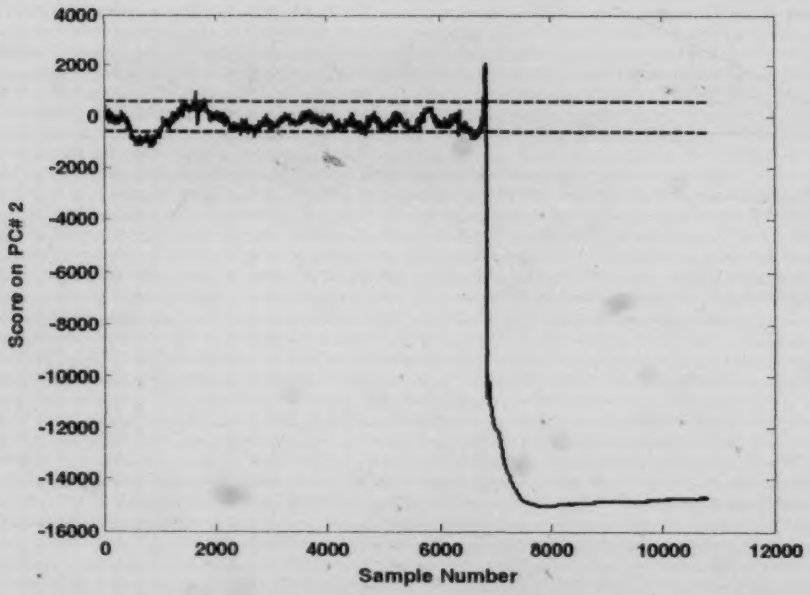


Figure 5-11(b): 2nd PC from validation data; this PC is used for monitoring

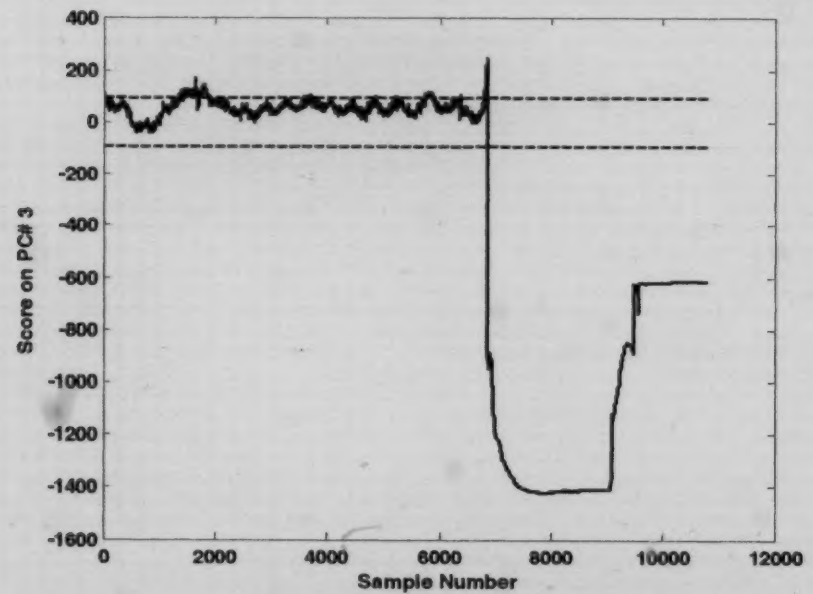


Figure 5-11(c): 3rd PC from validation data; this PC also gives the same result for monitoring decomposition, but the loading vectors sign and weightings seem unreasonable

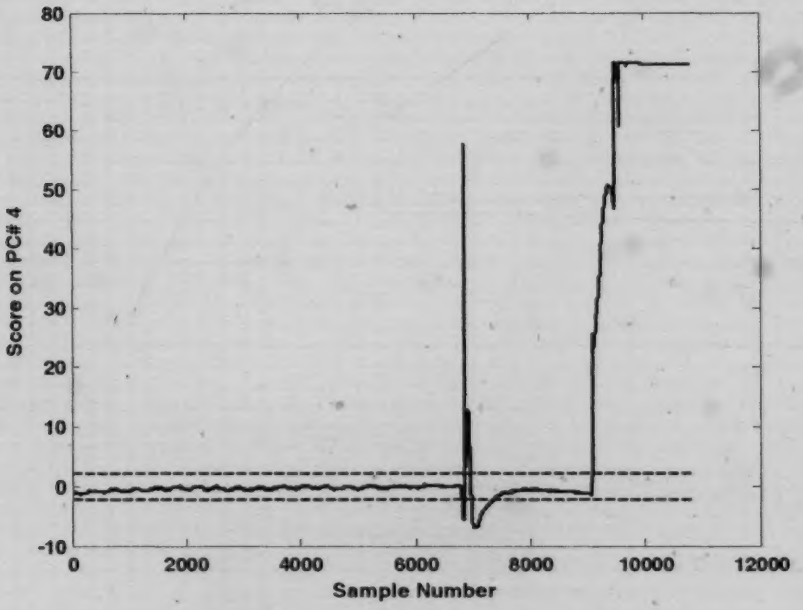


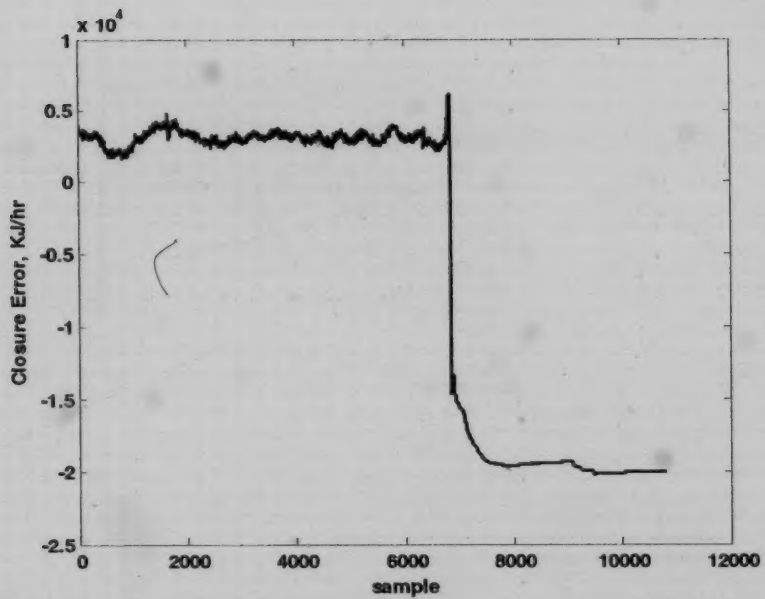
Figure 5-11(d): 4th PC from validation data

By comparing Figure 5-8, and Figure 5-11(b), we observe that the variation of error calculated from the theoretical model and the PCA model is exactly the same. These two figures are plotted again in Figure 5-12 and the similarity is clearly seen. This is another reason for selecting the 2nd PC as representative of the closure error. From the similarity in the above two cases, we may consider the mechanistic model (where approximate values of Cps are used) for monitoring and disregard the PCA model altogether. While this may seem feasible, the PCA model was chosen for the rest of the analysis for the following reasons:

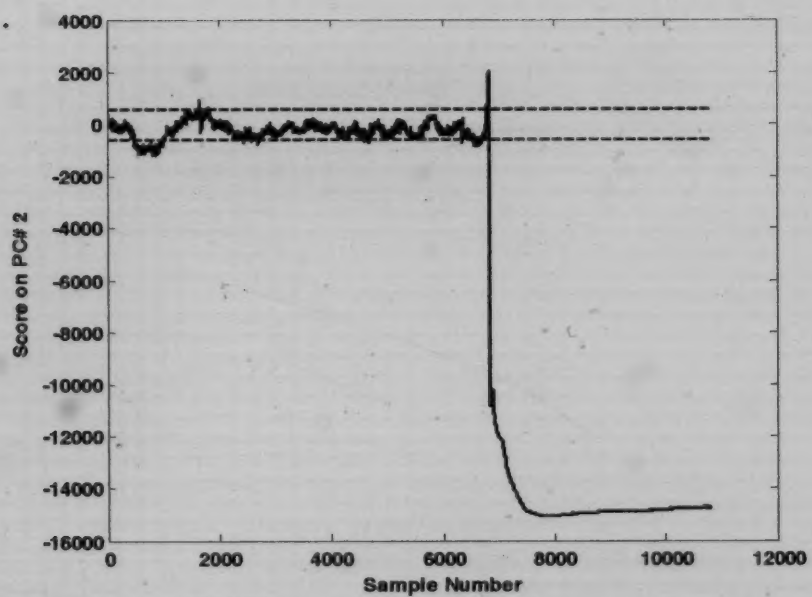
- The PCA model results in much smaller magnitude (orders of magnitude smaller) of the error compared to the mechanistic model. The thermodynamic property values used in the analytical expression are not known accurately and therefore this closure error has larger bounds. Projection of the data onto the loadings vector corresponding to the PC2 dominates the composite residual error and has

the correct signs. Therefore this vector is taken as a proxy vector of the energy balance closure error.

- In any case, the thermodynamic property values are likely to change with different grades. Since an important objective of this work is to monitor the reactor under different operating conditions and because it is possible to obtain the loadings vector in the residual space very easily from data sets in different operating regimes, the PCA based monitoring scheme has been found to be more practical and robust.



(a)



(b)

Figure 5-12(a): Error calculated from theoretical model (Error = $H_{react} - H_{mix} + H_{VA} + H_{ethy}$) assuming ΔH_R and C_p values are perfect; (b) Error calculated from PCA model (Error = $0.66H_{react} - 0.74H_{mix} + 0.0058H_{VA} + 0.078H_{ethy}$) assuming uncertain ΔH_R and Cp

As a final test to ensure the appropriate selection of PC, we looked at the squared prediction error (SPE), which is the linear combination of all errors resulting from PCs that span the null space. When one PC is retained in the PCA model, SPE or the Q-statisic is estimated as follows:

$$SPE = \left(X - t_1 p_1^T\right)^T \left(X - t_1 p_1^T\right) \tag{5-30}$$

Figure 5-13(a) shows SPE calculated for the validation data set. The variation of SPE is found to be quite similar to PC2 (see Figure 5-11(b)). Figure 5-13(b) shows a closure view of the SPE plot near decomposition. SPE starts to increase sharply at sample number 6780, the same time when PC2 (which is taken as a representative of the energy balance closure error) starts to increase. Therefore, monitoring SPE with a 1-PC model gives identical result as monitoring PC2 alone. Figure 5-13(c) shows a zoom in plot of PC2. This feature was checked for each of the case studies presented in this paper and similar result was obtained.

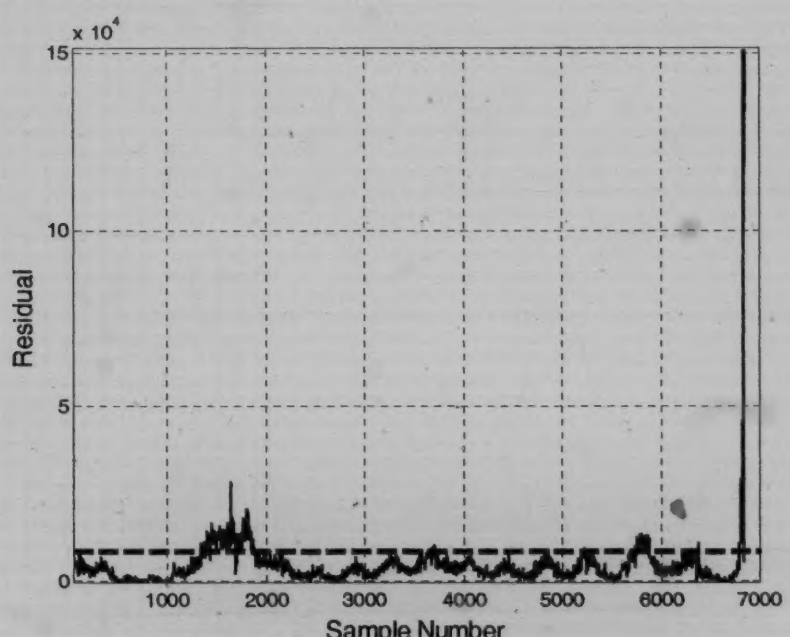


Figure 5-13(a): Squared Prediction Error (SPE) calculated from the PCA model

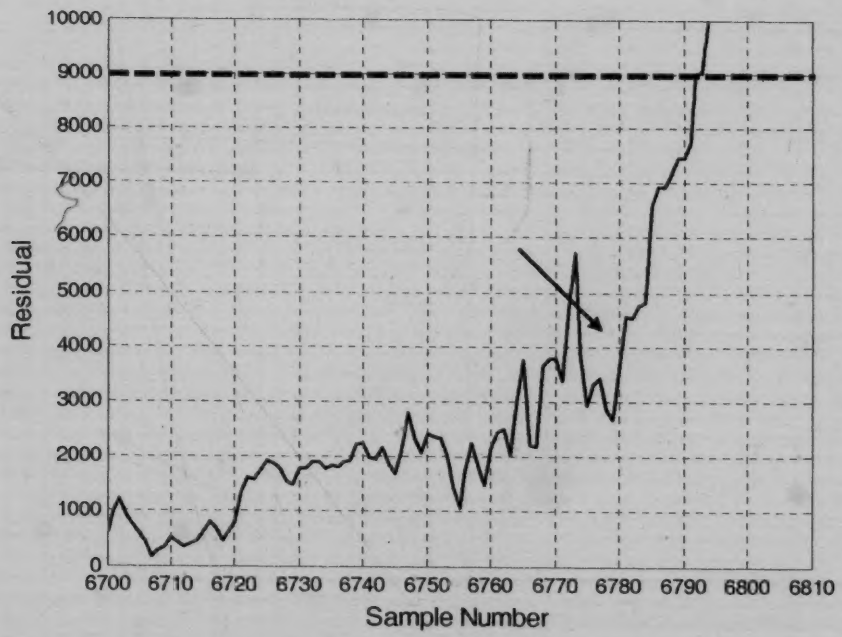


Figure 5-13(b): Zoom in of the SPE plot where the decomposition occurs

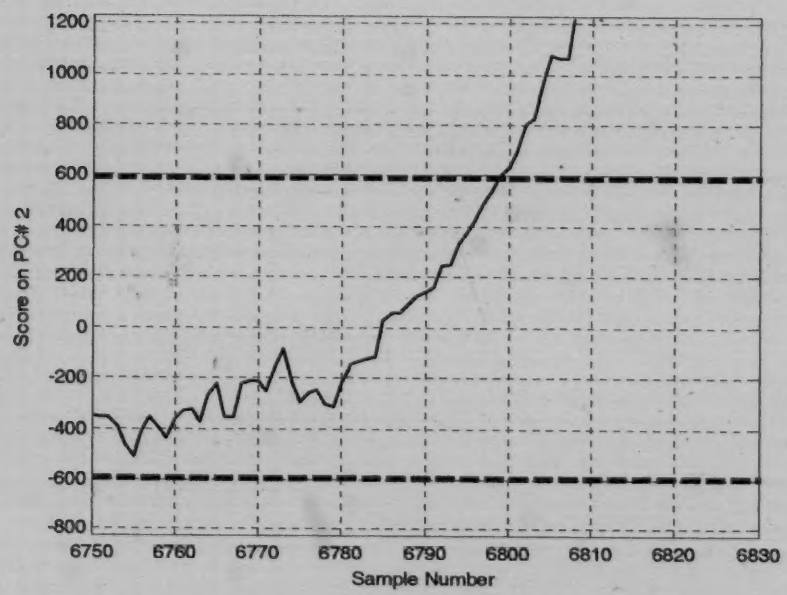
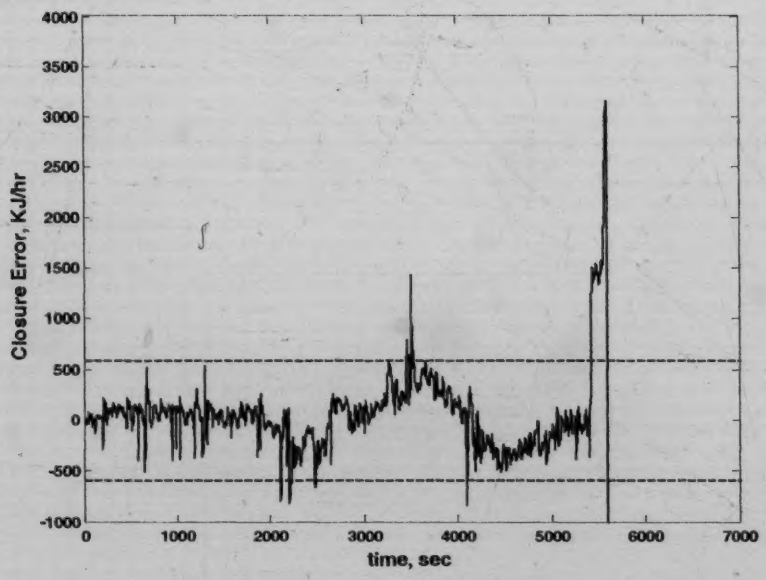


Figure 5-13(c): Zoom in of PC2 for validation data near the fault

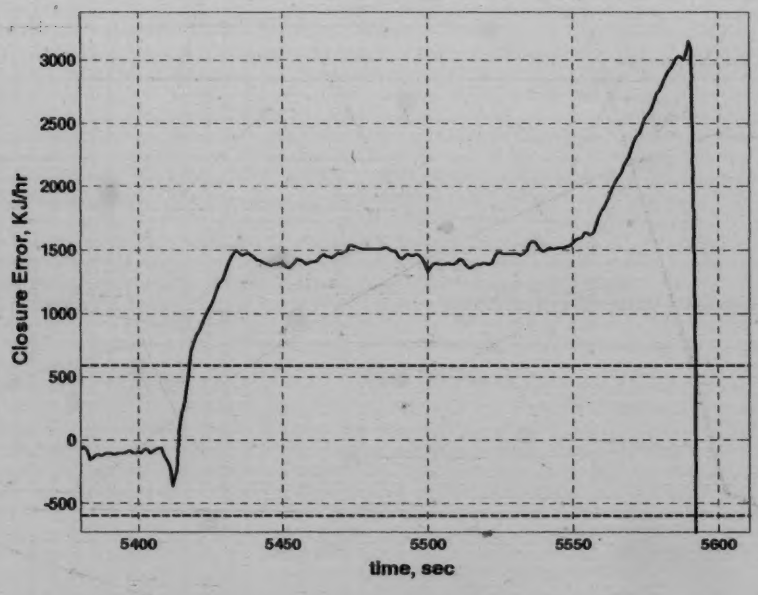
5.9 Results and Discussion

Once the model was identified, the next step was to use the model for detecting abnormal situations. Four sets of data, each of which contained some known faults, were used for validating the model. In the first two data sets, decomposition took place, followed by reactor shutdown. In the other two sets, the reactor temperature showed abnormal behavior which was similar to situations where decomposition took place. However, in these cases, the reactor recovered through operator intervention without any serious consequences.

Figure 5-14(a) shows the energy balance closure error from the first validation data set and Figure 5-14(b) shows a zoom-in of the same data near the fault. The data set contained 10800 samples of several measurements from the reactor and was collected using one second sampling interval. Decomposition took place near sample #5590, known from the plant history. A closer look into Figure 5-14(b) shows that at sample number around 5420 (170 seconds before the fault), heat balance closure error exceeded the safety limits and continued to increase until decomposition took place. Careful investigation of all important variables around the reactor revealed that ethylene conversion in the reactor suddenly increased after sample #5420, almost three minutes prior to the accident (Figure 5-14(c)). However, reactor temperatures as recorded by the thermocouples remained constant even though ethylene conversion began to change. After sample #5550, about 50 seconds before the fault, the lower zone temperatures showed an abnormal decreasing trend (Figure 5-14(d)). At the same time, the initiator flow rates were increased in order to compensate for the observed change (Figure 5-14(e)). Mid zone reactor temperature rapidly increased 40 seconds after this corrective measure was taken and resulted in decomposition at sampling time 5590. The sudden increase in ethylene conversion in the reactor resulted in an increase in heat generation which was not obvious from monitoring the reactor temperature records. The closure error, which shows the combined effect of heating and cooling inside the reactor was able to detect the change with reasonable lead time. Discussion with plant personnel suggests that three minutes would have given the operators sufficient time to bring an emergency reactor shutdown in effect, thus the decomposition could have been avoided.



(a)



(b)

Figure 5-14: (a) Energy balance closure error estimated from the PCA model for case study 1; (b) zoomed in region where fault occured

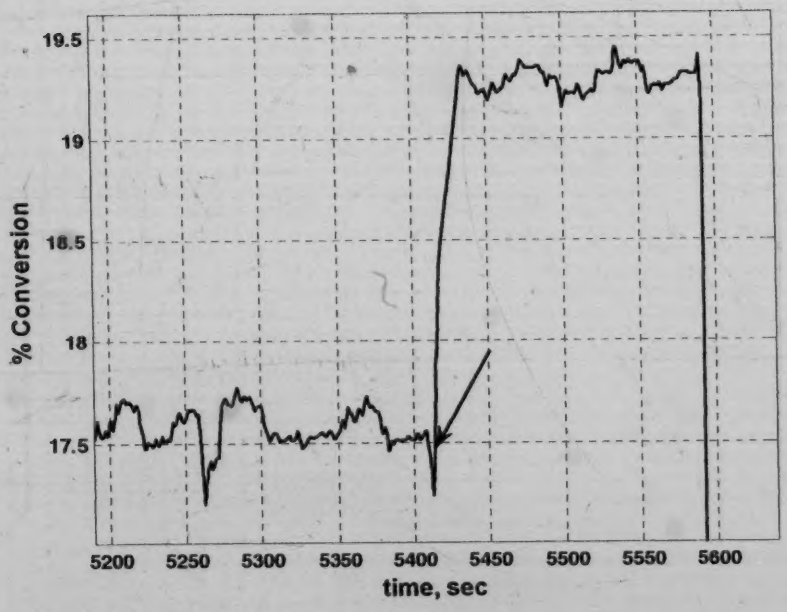


Figure 5-14(c): Ethylene Conversion during decomposition (Case study 1)

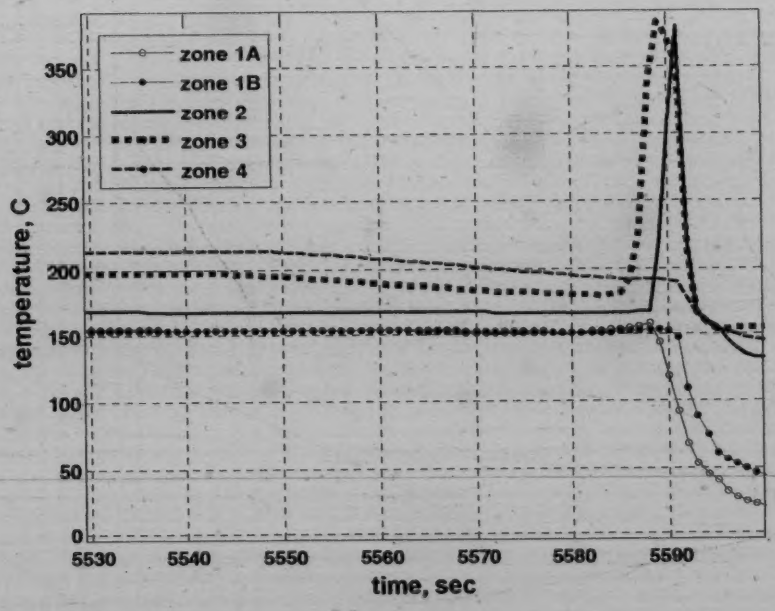


Figure 5-14(d): Reactor temperature profile during decomposition (Case study 1)

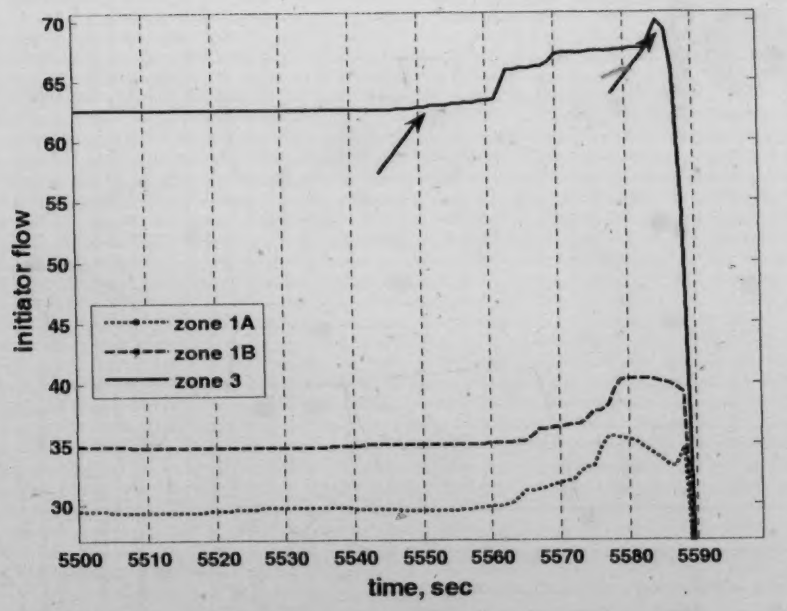


Figure 5-14(e): Initiator Flow rates during decomposition (Case study 1)

In the second case study, decomposition took place in a very similar manner during production of the same grade of EVA copolymer as in case study 1. The data matrix contained 8000 samples collected with a one second sampling interval. Decomposition occurred during steady production of the polymer at approximately sample #6834. Figure 5-15(a) and 5-15(b) show the estimated closure error from the model for this data set. The error started to increase at sample #6780 and went beyond the safety limit after sample #6800, which was 34 seconds before the decomposition. In this case, the reactor temperatures were among the only variables which showed abnormality. The lower zone temperatures (Figure 5-15(c)) started to decrease almost 42 seconds before the accident. The operators increased the initiator flow rate (Figure 5-15(d)) approximately at the same time and the final result was decomposition.